2819

COHESION IN THE SENTENCE

Its use in evaluating grammars

E.D.J. Schils

COHESION IN THE SENTENCE

Its use in evaluating grammars

E.D.J. Schils

PROMOTOR :

Prof.dr. W.J.M. Levelt

COHESION IN THE SENTENCE

Its use in evaluating grammars

PROEFSCHRIFT

ter verkrijging van de graad van doctor in de letteren
aan de Katholieke Universiteit te Nijmegen, op gezag
van de Rector Magnificus Prof.dr. J.H.G.I. Giesbers,
volgens besluit van het College van Dekanen in het
openbaar te verdedigen op donderdag 14 april 1983,
des namiddags te 4.00 uur

door

Erich Denis Johan Schils
geboren te Breda

Druk: Sneldruk Boulevard Enschede

1983

ACKNOWLEDGEMENTS

For the greater part the research reported in this study was performed while I was a staff member of the Mathematical Psychology department of the Catholic University of Nijmegen. I am endebted to those colleagues and students who aided me during this period with their valuable criticism and advice, their assistance in conducting experiments, their help in writing computer programmes when problems exceeded the capacities of a non-professional, and their participation as experimental subjects.

Much of the data analysis and preparation for this study was done in my present place of employment, the Faculty of Letters of this university. In this phase of the battle many difficulties were overcome thanks to the computer-technical assistance of Mr. F. Reelick, the organizational talents and moral support of Mr. H. Bours, and the advice of other colleagues whom I could consult with for various kinds of problems.

Finally I wish to express my gratitude to those who found themselves confronted with the task of wading through the many pages of this thesis in some way or another. These heroes must, alas, remain anonymous for various reasons.

Aan mijn moeder,

Aan Frieda, Fanny, Floortje en Peter

CONTENTS

CHAPTER 1 / SYNTACTIC COHESION AND THE INTERPRETATION PROBLEM IN LINGUISTICS

1.1 INTRODUCTION

Page 101 of Chomsky's Aspects of the Theory of Syntax (1965) contains a passage which is remarkable from a methodological point of view. In considering the sentence *he decided on the boat on the train* the author observes a greater degree of cohesion (his term) between *decided* and the prepositional object *on the boat* than between *decided* and the place adverbial *on the train*. Chomsky mentions this observation together with his decision to modify the rules of the base grammar. Instead of the old set of rules that -aside from irrelevant details- would assign to the sentence the structure of Figure 1.1., Chomsky chooses a set of rules that results in

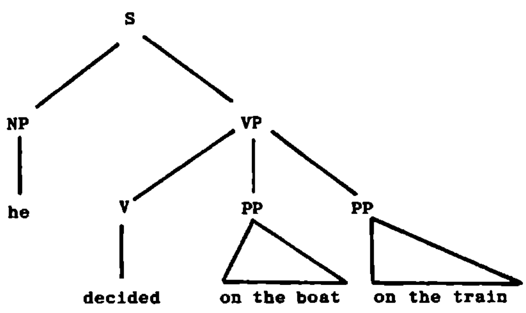


Figure 1.1 (S: sentence, NP: noun phrase, VP: verb phrase V: verb, PP: prepositional phrase)

the structure of Figure 1.2.
What is methodologically remarkable about this passage is the indeterminate status of the reported observation. Chomsky gives no explicit account of the way judgments on the cohesion between words or constituents should be related to the formal properties of hypothesized sentence structures. Furthermore, nowhere in his book is it announced that he considers cohesion judgments to be relevant facts that should be accounted for by a grammar. In view of these omissions it cannot be determined whether Chomsky's cohe-

sion observation is in support of his theoretical decisions.

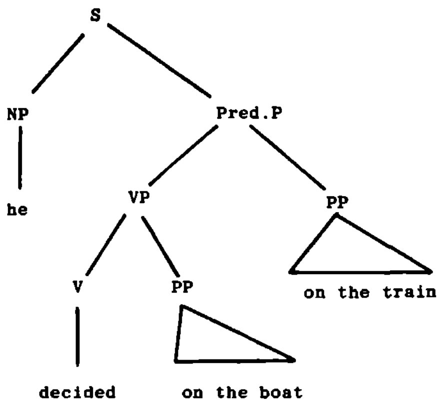


Figure 1.2 (Pred.P: predicate phrase)

Let us turn to the question: are cohesion intuitions to be regarded as relevant linguistic facts? On the one side we could argue that the mere fact that Chomsky and other linguists have occasionally used cohesion arguments (see Section 2.2.1) is sufficient to assume that they are. On the other, cohesion arguments do not occupy a prominent place in syntactic theorizing (e.g. Aspects contains only one isolated example involving an observation of cohesion) and this might be regarded as an argument against their relevance. The above passage would then reduce to a simple inconsequential heuristic aside.

If, for the moment, we assume that the answer to the above question is affirmative, then we still do not know whether cohesion observations confirm Chomsky's theoretical decision. This will of course depend on the answer to the other neglected question: what is the exact nature of the correspondence between the hypothesized structure and a cohesion judgment? Levelt (1974c) has made an attempt to make explicit what he assumes Chomsky performs implicitly in the above mentioned passage. To Levelt, Chomsky handles a rule of correspondence involving an inverse relation between the cohesion of -let us say- two linguistic entities x and y, r(x,y), and the height of the lowest node dominating both x and y: the higher that lowest dominating node, the lower the degree of cohesion and vice versa. According to this rule of correspondence, the greater cohesion between *decided* and *on the boat* that that between *decided* and *on the train* is related to the fact that the first pair is dominated by a lower node than the second pair. But, obviously, this rule of correspondence would result in implausible predictions in respect of cohesion judgments given for pairwise comparisons in which the highly dominated but intuitively strongly related pair (he,decided) is involved. For these pairwise comparisons, application of the above rule to the structure of Figure 1.2 is likely to underpredict the cohesion between *he* and *decided*. In all probability, the pair (he, decided) will be judged more cohesive than the pair (on the boat, on the train), for instance, and the inverse prediction will be violated. Of course this argument could be refuted by rejecting Levelt's interpretation as an explicit account of Chomsky's implicit intention. Alternatively, one might resort to the argument already mentioned in connection with the first question: cohesion observations are not actually among the sort of facts which Chomsky and other linguists take into account in constructing their theories. These defences, however, would only accentuate the fact that in this passage of Chomsky's Aspects too much has been left implicit for a thorough understanding.

There is, so to speak, a considerable gap between formal syntactic theory on the one side and intuited cohesion on the other, and it is far from clear how this gap could be bridged within the purview of current linguistic methodology. This brings us to the very reasons for opening the present study with a quotation of the above passage from Aspects. Firstly, it was the intention to exemplify the conviction that, for some areas of linguistic performance at least, the relation between linguistic theory and linguistic fact is itself in need of careful study. Secondly, it is intended to introduce this relation, by example, as the general concern of this book. Thirdly, the example should serve to concretize the triad of themes constituting the specific concern of this study: syntactic structures, cohesion judgments, their interrelations and correspondence. The way in which these themes come in for study in this book is intended (i) as a specific exercise in a more general linguistic and methodological approach to the problem of the relation between syntax and observable data, to be discussed further in this introductory chapter; (ii) as a suggestion for other studies which, from the same general perspective, would try to examine the relation between syntactic structures and other linguistic facts.

1.2 *METHODOLOGICAL AMBIGUITIES IN THE USE OF COHESION*

The cohesion observation referred to in Section 1.1 revealed a methodologically equivocal character, manifesting itself as a gap between formal syntactic theory and empirical data. We said that current transformational methodology can hardly be expected to bridge this gap. Part of the problem concerns the theoretical side, another part relates to empirical considerations.

As for the *empirical* side, one should realize that cohesion intuitions *are data of linguistic performance. As such, they, like other instances of* linguistic performance (e.g. the phenomena of speech production, perception, retention, acceptability judgments, paraphrase judgments), are to be conceived of as indirect manifestations of linguistic competence. The latter comprises what grammar purports to be a theory about. The former, the indirect reflections of an underlying competence in various instances of linguistic performance, must somehow yield the theory's empirical foundation. The distinction is not unlike that made in the methodology of the behavioural sciences between the genotypic and phenotypic levels of behaviour. Coombs (1953, p. 488-9), for instance, uses the following expression: "The phenotypic level refers to the observed or manifest behaviour; the genotypic level to an inferred, hypothetical, latent level of behaviour underlying or generating the phenotypic level". And some lines further he writes: "The manifest behaviour is implicitly regarded as a function of the individual's genotypic ability and certain characteristics of the stimulus situation". In the "methodological preliminaries" to Chomsky's Aspects the distinction takes the following form:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech community, who knows its language *perfectly and is unaffected by such grammatically irrelevant conditions* as memory limitations, distractions, shifts of attention and interest and errors (random or characteristic) in applying his knowledge of the language in actual performance. (Aspects, 1965, p. 3).

In this vein, performance data can be regarded as manifestations of competence and extraneous, i.e. non-grammatical, factors. If, therefore, one is to develop a theory of linguistic performance, rather than a theory of competence, the performance data (D) will have to be specified as some function (F) of a grammar (G), together with extraneous factors of a systematic $(e_1, e_2, \ldots, e_k)$ and/or random nature (ε). Symbolically,

$$D = F(G; e_1, e_2, \ldots, e_k; \varepsilon) \tag{1.1}$$

This symbolism should be regarded as a general and versatile frame of reference. There are many instances of performance data and the possibility should not be excluded that D's can be obtained (by selection, manipulation or control) that allow for a zero or near-to-zero option with respect to the e's and the ε. In its general form, the symbolism clearly reflects the fact that D and G "are not on a par": G is not a theory of D. Hence part of the gap seems to be due to the paradoxical position of the linguist who finds himself having to test his theory against data which it is not about. Or, as an operationalist would formulate it: cohesion judgments are invalid and/or unreliable indicators of underlying competence.

As was already mentioned, other problems arise on the *theoretical* side of the gap. From Chomsky's cohesion observation it was impossible to decide whether or not the structural modification depicted on the first page of this study was supported. This is due to a deficiency of theory and has no bearing on the "quality" of the data. What is lacking is a theoretical supplement to the syntactic theory, specifying the correspondence between the formal properties of syntactic structures and the required structure of the data of cohesion. Without such a theoretical annexe it is impossible to derive the empirical consequences of structural decisions, not even for the ideal, genotypic case. In other words, such a supplement is indispensible for testing purposes. This not only applies to the area of cohesion phenomena, but to other instances of performance data as well.

1.3 REMEDIES FOR THE METHODOLOGICAL AMBIGUITIES

After this digression into some aspects of the methodological distance between formal syntax and cohesion data, we are now ready to anticipate the content of this thesis. As a general characterization, the study may be said to deal with the problem of bridging the gap here indicated. More specifically, concrete proposals for achieving this end will be advanced and elaborated. In the light of the foregoing digression it will be clear that these proposals will imply measures (i) for curing the theoretical deficiency signalized above; (ii) for coping with the extraneous factors of the judgment process.

1.3.1 Measures with respect to the theoretical deficiency

The interpretation problem. The remedy for the theoretical deficiency has already been suggested above. This suggestion is not new: it is due to

Levelt (1974b,c), who has explicitly paid great attention to the *relation* between theory and data in linguistic research. He calls this problem the *interpretation problem*. His general proposal for those instances of the interpretation problem where current tranformational methodology falls short, amounts to the construction of an *interpretation theory*. This is the already suggested theoretical supplement to syntactic or linguistic theory, mediating between formal theory and the data. Levelt argues that this interpretation theory ought play the same role in linguistics as measurement theory does in the social sciences. In terms of the symbolism given above, it has to specify the nature of the function F. Levelt elaborates his general methodological point of view for two areas of linguistic performance: acceptability and cohesion.

Clearly, the latter is of particular concern to the present study and will actually comprise its very starting point. Levelt demonstrates how interpretation theories can be constructed both for the constituent grammar and for the dependency grammar (see Chapter 2). He gives an explicit account of *which* formal properties are considered to determine cohesion judgments and *how* the latter are held to be related to the former. In essence, this formalization adopts a distance metric over the syntactic structures and inversely relates the cohesion between syntactic entities to their distance. This enables the derivation of testable predictions with regard to the cohesion observations from the formal linguistic structure. Levelt's cohesion judgment research is thus both an exercise in linguistic interpretation and a comparison of the structural adequacies of the constituent and dependency models. The tentative conclusions drawn from Levelt's investigations, to be considered in detail in Chapter 2, are in favour of the dependency grammar.

The comparison of these formalisms is continued in this study, where the conclusions tend to point in the same direction. The argumentation adopted however, is different. In the final section of Chapter 2 arguments are put forward for the abandonment of the entire family of distance models, to which Levelt's belong, for the description of cohesion. Alternative interpretation theories of an essentially different nature will be presented in Chapter 4.

Another problem which Levelt's cohesion models share with our alternative versions introduced in Chapter 4, is their vulnerability resulting from the severe deterministic formulation of their interpretation theories in terms of the ideal or genotypic case. Phenotypically, however, the data are clearly of a non-deterministic nature; moreover, they give a strong impression of being co-determined by extraneous, non-grammatical factors. The deterministic formulation in terms of the ideal case is therefore bound to be rejected, for it makes syntax and syntax alone responsible for all of the variation in the data, part of which must however be considered as non-syntactic. The alternative interpretation theories are accordingly further elaborated and adjusted. In Chapter 5 probabilistic versions supercede the deterministic models. In subsequent sections these are adjusted so as to cope with some extraneous factors of the judgment process. This brings us to the second kind of measure implemented for bridging the gap between formal syntax and cohesion data.

1.3.2 Measures with respect to the "extraneous factors"

The extraneous factors co-determining performance data naturally complicate the endeavour "to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer (Aspects, p. 4)". Expressed differently: they complicate the testing of a competence theory. This aspect of the "methodological gap" still awaits resolution even after the adoption of an interpretation theory in the sense of the *preceding section which translates the formal aspects of syntactic theory* into ideal genotypic behaviour. Let us consider an elementary example. The simplest possible interpretation theory able to mediate between a transformational grammar, G, and judgments of acceptability with respect to strings of words, s, would read:

(1.2) s is acceptable $\Leftrightarrow$ $s \in L(G)$ i.e. s is grammatical
(where L(G) denotes the language, generated by G).

But it is widely known that things are not that simple; should the reader doubt this let him turn to Greenbaum (1976) where many instances can be found of extraneous factors co-determining acceptability.

This methodologically equivocal state of affairs is depicted in a simplified and schematic way in Figure 1.3. The two boxes with solid outlines represent the different sides of the methodological gap under discussion: competence theory (G) and performance data (D). Their unequal sizes reflect the fact that they are not on a par: the former is not a theory of the latter. The "real world" counterpart of grammar is represented by the equally large dotted square labeled "competence". This accentuates the fact that competence is conceived of as a mental reality rather than a mere theoretical construct. The theoretical counterpart of performance data would be a performance theory, and this is represented by the large dotted square to the left of the diagram. In conformity with the foregoing, performance

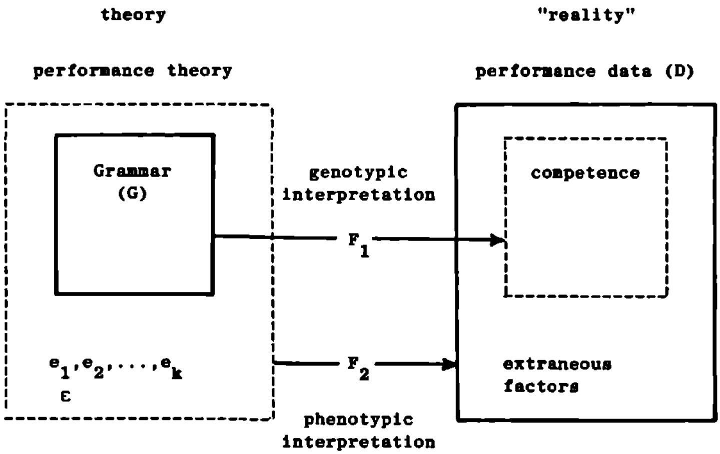


Figure 1.3 Schematic representation of the methodological gap between COMPETENCE THEORY and PERFORMANCE DATA

data are depicted as the resultant of competence plus extraneous factors, the theoretical counterparts of which are the e's and ε. The arrow F_1 pointing from *grammar* to *competence* represents the interpretation theory as implied by the preceding section. It is clear that the distance to performance data is only partially bridged, in that it relates only to the ideal, genotypic behaviour.

In terms of the figure, there are two ways in which the solid outline boxes could be made equal. On the right the solid box could be reduced to the size of the dotted one, on the left, enlarged to the size of the dotted one. Such modifications to the figure reflect corresponding measures for bridging the residual methodological gap. The modification on the right, or empirical side, would correspond with measures for somehow minimizing the extraneous factors, so as to obtain the best possible direct reflection of competence. Modification of the left, or theoretical side, would correspond with the idea that a grammar's adequacy can also be tested by demonstrating its indispensibility as a component of a succesful performance theory. This would imply a modification of the theory and a corresponding extension of the interpretation theory (in the figure: the arrow, labeled F_2). This latter point of view is the one which we shall adopt in the present investigation of cohesion. It has much affinity with a relatively recent formulation of the objectives of grammatical inquiry by Joan Bresnan (1978), to be discussed later in this chapter.

Current transformational methodology. In our view, the former approach comes close to what current transformational methodology, henceforth CTM, is doing. Typical in this connection is CTM's adoption of the linguist's own intuitions of acceptability as the empirical basis for the notion of grammaticality. The reasons for this resort to the linguist's introspection in a problem which was originally characterized as "determining from the data of performance the underlying competence" are widely known. Resort to acceptability intuitions rather than occurrence in a corpus of recorded speech, for instance, bypasses those extraneous factors which either exclude grammatical or include ungrammatical sentences within such a corpus. Resort to the linguist's rather than the naive native speaker's intuitions should reduce those extraneous causes of invalidity or unreliability referred to in what is sometimes called the "wine-taster's argument". There are many linguistic distinctions whose discernment requires so much training that it is hardly surprising they elude the naive informant.

Similarly, introspection is preferred to data gathering procedures of various kinds involving naive informants. This will become clear from a small review of quotations, all taken from Aspects (p. 19):

> "... no adequate formalizable techniques are known for obtaining reliable information concerning the facts of linguistic structure...", "there are (...) very few reliable experimental or data-processing procedures for obtaining significant information concerning the linguistic intuition of the native speaker", "... when an operational procedure is proposed, it must be tested for adequacy (...) by measuring it against the standard provided by the tacit knowledge that it attempts to specify and describe", "allusions to presumably well known procedures of elicitating criteria or objective methods simply obscure the actual situation in which linguistic work must, for the present, proceed", "... there is no

reason to expect that reliable operational criteria for the deeper and more important theoretical notions of linguistics (such as "grammaticalness" and "paraphrase") will ever be forthcoming."

All these quotations strongly suggest that what Chomsky resists -rightly we think- is in fact a naive *operationalistic* use of tests, experimental data, judgment tasks and elicitation procedures as *aprioristic* indicators of linguistic notions such as grammaticality, ambiguity, paraphrase and so on. He resists an elaboration of correspondence between theory and data in the form of simple operational definitions, equating "grammatical" with "what the native speaker accepts" or "ambiguous" with "what a native speaker in his performance demonstrates to interpret in more than one sense". This would again allow extraneous factors of various sorts to invalidate the data.

In what he formulates as "a dilemma between objectivity and insight" Chomsky (op.cit., p.20), evidently, chooses for insight. In this connection, however, it is worth anticipating our own employment of elicitation procedures, namely, for cohesion, which is not operationalistic at all. In the present study, relatedness judgments do not comprise an operational definition of something like syntactic coherence. They are regarded as data to be analyzed and whose explanation requires, among other things, the notion of syntactic coherence. This is, of course, a quite different attitude.

CTM's resort to introspection is naturally not without risks. It leads linguistic practice into a situation in which the selfsame person is often the source of both theory and data; in which inspection of the data to suggest ideas is seldom discernable from viewing the data to test the ideas. In the effort to preserve objectivity of research it is therefore required that only *clear cases* should provide the basis for description and explanation, i.e. those linguistic facts which are intuited in consensus by professional linguists.

CTM and cohesion. Before addressing ourselves to the second means of spanning the methodological chasm between competence theory and performance data, we have to consider the reasons which make it necessary to seek an alternative to CTM. Prior to these considerations it should be remarked that the two approaches are here regarded as complementary rather than competetive. CTM has proved to be a rewarding methodology, enabling rapid development of the study of grammar. It rightly and succesfully takes advantage of a circumstance specific to the study of linguistics: in principle, linguists who are native speakers permanently carry the relevant data with them. Acceptability observations can be introspectively performed at any time and place during the process of inquiry. Unlike observations of "anxiety", "drive", "agression" of "likedness" and so on, they can be written down self-referentially on paper in the form of starred and un-starred strings like **the went girls* and *the girls went*. They are thus easily exchanged and their "clearness" checked by testing whether the consensus requirement is met. Moreover, for a long period the state of the evidence in linguistics was characterized by the availability of a mass of unquestionable data. Since no adequate description, let alone an explanation of these data was available, theory construction took priority over considerations of data collecting methods.

Nevertheless, CTM has given occasion for criticisms of various kinds, especially in connection with the resort to introspective judgments. It will be worthwhile to pay attention to this criticism since some of its aspects concern us here: Cohesion intuitions result from introspection, albeit the introspection of the naive informant rather than that of the linguist. We shall proceed by taking a look at Labov's (1975) "What is a linguistic fact?" because many of the more important issues raised in connection with the use of introspection in linguistics can be found in this publication. In a nutshell, Labov's general attitude towards the use of introspective judgments can be characterized as reserved. The author is willing to accept them as linguistic facts provided they represent cleas cases. But he harbours a suspicion against *variation* in introspective judgments. Labov (op.cit.) reviews certain investigations in which variation of introspective judgments could be criticized as "an ominous sign of idiosyncratic and extraneous influences on linguistic data" (p. 41). His reaction is to abandon, for the time being, variation in introspective data as non-linguistic facts. Where the clear case principle fails to guide the study of introspection, one should rather study the more reliable facts of language use. Naturally enough, we feel challenged by this position. One needs but little experience of the phenomena of cohesion for recognizing their variable nature both within and over informants. Must it necessarily follow that cohesion be excluded from the facts of linguistics? Certainly not any more than body weight should be excluded from biometric statistics merely because people have different weights in the morning and evening.

In this connection one should realize that the use of introspective judgments can be criticized on two levels. This can be clarified by the introduction of a distinction made by Kaplan (1964), between the *logic-in-use* and the *reconstructed logic* employed in the pursuit of a science; in essence the distinction between current research practice and its idealized description. The reconstructed logic of transformational research, for instance, uses clear cases as a basis for the description and explanation of the language; the logic-in-use tries to realize this ideal in actual research. The criticism of introspective linguistic research is partly a criticism of its logic-in-use, partly a criticism of its reconstructed logic. In the first case it applies to situations in which the linguist is unfaithful to his working principles, in the second it concerns these working principles themselves. Two examples, taken from Labov's section on the "wholesale rejection of linguists' judgments" (p. 14 ff.), may serve as an illustration.

Criticism of the logic-in-use is exemplified by a study by Nancy Spencer (1973) referred to by Labov. From this study emerged a considerable disagreement among both naive subjects and linguists concerning the grammaticality of 150 sentences which have been used for purposes of grammatical argumentation in studies by Perlmutter, Carlotta Smith, Postal, Ross, Rosenbaum and Lakoff. Obviously, in these studies the clear case principle has not been obeyed, or, as Labov puts it, the logic of linguistic inquiry there employed had been to assume consensus rather than test it. Whether the lack of consensus thus established forms a sufficient reason for the "wholesale rejection of linguists' judgments" is to be doubted. Such criticism might be better interpreted as an exhortation to linguists to remain true to their working principles.

The reconstructed logic itselfs is involved in Labov's discussion of a study by Grinder and Postal (1971). In that study the sentence *John didn't leave until midnight, but Bill did* is regarded as an experimentum crucis for the resolution of the debate between generative semantics and the interpretivist theories of syntax. The authors themselves reject the sentence as ungrammatical, Chomsky accepts it and varying responses are given by linguists and other speakers consulted by Labov. An essential deviation from the reconstructed logic of transformational research is Grinder and Postal's reaction to the lacking consensus. Instead of abandoning the disputed example as an unclear case, they argue that the varying judgments are to be considered as dialect differences. Moreover, they regard their own theory as superior, since it can explain these differences by deducing them from different applications of rules or constraints in the grammar. The authors consequently argue that the facts of variation are not to be ignored; they provide the opportunity for a stronger test of the explanatory adequacy of a syntactic theory than the facts of a single dialect only.

The important point we wish to bring out in referring to these examples is that it is impossible to decide independent of theory whether something like the clear case principle is a valuable working principle for a particular logic of inquiry. The principle is in line with a theory such as standard transformational grammar claiming to be about a homogeneous language community whose members genotypically operate under the same set of deterministic rules and constraints. In this situation the clear case principle is to a certain degree intended to bypass the grammatically irrelevant factors held responsible for the phenotypic inter-individual variation. For a theory, however, which regards phenotypic variation as a reflection of genotypic, e.g. dialect differences (see, for instance, Elliot, Legum, Thompson, 1969; Greenbaum, 1973; Legum, Elliot, Thompson, 1974), the principle stands at least in need of revision. And finally, the principle is doubtless alien to a theory which regards phenotypic variations as a reflection of genotypic stochasticity (Carden, 1973; Sankoff, Rousseau, 1973). Such a theory invites an essentially different methodology.

The clear case principle is, in other words, a crucial issue for CTM only, and its frequent violation in transformational research should thus be taken seriously. But the principle can never be an absolute decree governing the study of introspective judgments as such. That introspective data such as cohesion judgments exhibit variation is in itself no valid argument for excluding them from the domain of linguistic facts. Neither are we ready to embrace such a *general* exclusion on the basis of Labov' review of *particular* investigations in which systematic variation in introspection seemed to be an artifact of poor elicitation procedures rather than a reflection of putative dialect differences.

As already stated, the study of variation in linguistic intuitions necessarily demands the revision or abandonment of the clear case principle. In CTM, however, the clear case principle serves as a sort of security measure in a practice where the selfsame person may generate both theory and data, and in which heuristics and testing are often intertwined. Rejection of the clear case principle, therefore, increases the urgency of disentangling these matters and encourages adoption of the methodology of behavioural sciences.

1.3.3 The realistic approach

The second means for spanning the methodological gap between competence theory and performance data takes a different position with respect to extraneous factors, the e's and the ε in Formula (1.1). In cases where these factors can not be eliminated or even reduced by methodological measures (selection, manipulation, control) but seem to be inherent aspects of the data, the situation might be accepted as it is. The extraneous factors are not then to be considered as "bad qualities of the data", but as properties to be explained. And the data of performance which are affected by these factors are not then to be discarded as "unreliable" or "invalid" indicators of underlying linguistic structure, but rather the objective of a theory of performance. This, however, is not incompatible with the above-mentioned goal: the development of a competence theory. It means only that another approach is taken. The question of a given grammar's adequacy becomes the question of whether or not it is an indispensible component of a successful integrated performance theory.

The idea that "a reasonable model of language use will incorporate, as a basic component, the generative grammar that expresses the speaker-hearer's knowledge of language" is certainly not new. It can be found in this extract from Aspects (p.9); it can be found again in a relatively recent formulation of the objectives of transformational grammar by Joan Bresnan (1978). But what in the former publication seems to be a linguistically uncommitted imperative directed at investigators of language use, is in the latter a requirement having repercussions on the adequacy of the competence theory too. In this context Joan Bresnan formulates two basic research objectives for transformational grammar: the *characterization* problem and the *realization* problem. The first task is to characterize the grammar representing the language user's knowledge of the language; the second is to specify the relation of the grammar to that model of language use of which it is a part. Thus far, argues the authoress, linguists have been too exclusively concerned with characterizing grammar and have neglected its realization in models of language use. The realization problem has generally been delegated to psychologists. This delegation has led to rather pessimistic conclusions (see Fodor, Bever, Garret, 1974) with respect to the psychological realizability of various facets of transformational grammar in the sense of Aspects. It is Bresnan's opinion that the incorporability of a grammar in a model for language use offers as much an invitation to linguists as it does to psychologists. Of the two possible reactions to the negative results of realization research, namely, the psychologist's and the linguist's one failed to appear. Many psychologists have resorted to non-linguistic models of language use, and rightly so, in as far as the linguistic models turned out to be impracticable. For linguists, however, the non-realizability of their models is hardly something to which they can remain indifferent.

> "If distinct grammatical rules were not distinguished in a psychological model under some realization mapping, the grammatical distinctions would not be 'realized' in any form psychologically, and the grammar could not be said to represent the knowledge of the language user in any psychologically interesting sense." (Joan Bresnan, 1978, p. 3)

This state of affairs should induce linguists to modify their theories so as to cope with the shortcomings. According to this point of view, henceforth to be referred to as the *realistic* one, Joan Bresnan herself proposes notable modifications to the Aspects model. The realization problem is of just as much direct concern to the linguist, as is the characterization problem.

In accordance with these considerations, the further elaboration of our cohesion models, from Chapter 6 onward, will be in the direction of an *integrated performance theory*, in which non-syntactic determinants are explicitly incorporated. This elaboration can be conceived of as an extension of the *interpretation theory* in Levelt's sense (the F_1 in Figure 1.3) to an *auxiliary* theory in Blalock Jr.'s (1971) sense (F_2 in Figure 1.3). Blalock devotes a chapter to "the operationalism controversy" (also Chomsky's dilemma !) in sociology, under the significant title: "the Measurement Problem: a Gap between the Languages of Theory and Research". The author's proposal for resolving the dilemma amounts to the construction of an integrated theory made up of two sub-theories: a general or main theory and an auxiliary theory. The main theory comprises the basic ideas in general terms. The auxiliary theory is indispensible for testing purposes; it is specific to the research design, the population under study and measuring instruments. It is moreover the domain reserved for the explicit incorporation of disturbing variables.

Following these general lines it will be demonstrated in this study how a particular aspect of the characterization problem, the assignment of syntactic structures to sentences, can be *realized* (in Joan Bresnan's sense of the term) with regard to native speakers' intuitions about such structures and the way these are expressed in cohesion judgments. In this sense, this thesis attempts a demonstration of how the *structural adequacy of grammatical formalisms and grammars* can be tested by resort to cohesion judgments. The denotation "structural adequacy" is here preferred to the polysemic "descriptive adequacy". The latter term also refers to the weak generative capacity of the grammar: the generation of all and only the sentences of the language. "Structural adequacy" is intended to refer only to the strong generative task: the assignment of the "correct" structural descriptions. The formulation indicates that structural adequacy will be studied on two levels. In a broad sense of the term it is the choice between grammatical formalisms which is at issue, viz. the choice between the constituent and the dependency formalisms. In a narrower sense of the term it is the options within a formalism that are the subject of study. In this case the choices take that form shown in the example from Chomsky's Aspects at the beginning of this chapter.

In this connection, however, it ought to be remarked that our claims with respect to the structural adequacy of particular types of grammar should not be overestimated. The claims in respect of *content* are of secondary importance compared to what is the main objective of this thesis: to contribute to the *methodology* of linguistic research. As far as content is concerned, the study is merely a close up of the comparison between the constituent and dependency formalisms in which decisive arguments for a choice between them will be sought in vain.

The intention of this thesis has been to use cohesion judgments as a frame of reference for the elaboration of practical methodological consequences based upon the general methodological point of view outlined in the present section. The hope is thereby entertained that at least some aspects of this exercise in linguistic interpretation, as demonstrated for a specific realization problem, will turn out to be useful for linguists wishing to approach other realization problems from the same general perspective.

CHAPTER 2 / THE STATUS QUO IN COHESION RESEARCH

This chapter will be devoted to a closer specification of the triplet of themes which are the concern of this study: syntactic structures, cohesion judgments and the problem of interpreting the former in terms of the latter. In the first section we will specify syntactic notions as far as they are needed for a proper understanding of the current state of cohesion judgment research as represented in Levelt's (1974c) study. This study forms the starting point of our treatment. We will limit ourselves to a minimum of technical detail, and assume general knowledge of the notion of constituent structure, surface and deep structure of the Aspects' sense and the like.

Two subsequent sections of this chapter will deal with the particular kind of observations we are interested in: cohesion judgments. The relevance to linguistics, as well as certain procedures for collecting this kind of data will also be discussed.

Following these sections on syntactic structures and cohesion judgments we will be prepared to address ourselves to Levelt's interpretation theories. Both his constituent model and his dependency model for relatedness judgments will be dealt with. A final section containing a critical review of these interpretation theories will conclude this chapter.

2.1 CONSTITUENT STRUCTURES AND DEPENDENCY STRUCTURES

Formal definitions of constituent-theory and dependency-theory* can be found in many sources in the literature. We feel justified therefore in omitting a repetition of these definitions here and refer the interested reader to texts such as Chomsky (1963) for C-theory, Hays (1964) and Gaifman (1965) for D-theory or Levelt (1974a, b) for both. Grammars provided by either of the two theories perform a double task. Firstly, they generate strings over a vocabulary of terminal elements, thereby defining a language. Gaifman (op.cit.) demonstrated that C- and D-theories are equivalent in this respect: they generate the same classes of languages. Secondly, they assign structural descriptions to the strings generated. In this respect the theories

* Henceforth, in compounds with "constituent"- and "dependency"- abbreviations using "C-" and "D-" will often be used.

differ. For our purpose it is this difference which is of major importance: cohesion phenomena are essentially structural phenomena. Chapter 3 of this study will be devoted to a detailed formal comparison of these formalisms. In order to prepare for the discussion of Levelt's interpretation theories in Section 2.3 of the present chapter, an informal introduction through examples will suffice.

Both C-structures and D-structures can be regarded as rooted trees. These trees consist of labeled nodes connected by directed branches. The left-to-right structure of such trees can be captured by defining a relation "is to the left of" which may hold between nodes. The C- and D-structures, however, differ in the theoretical conceptions of sentence structure which are expressed by means of such trees. More specifically, different theoretical interpretations are to be attached to the nodes and the branches. Let us illustrate this by comparing the most famous museum piece of generative linguistics, the C-structure for the sentence *the man hit the ball* given in Figure 2.1 (a) with the D-structure displayed in Figure 2.1 (b).

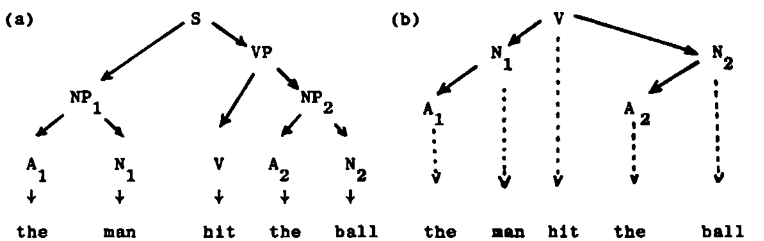


Figure 2.1 C-structure (a) and D-structure (b) for *the man hit the ball*

Both these structures can be regarded as the product of repeated applications of grammatical rules, conceived of as instructions for tree formation*.

In the case of the C-structure, the first instruction is to represent the starting symbol of the grammar, S, as the initial S-labeled node (the root) of a tree i.s.n. (read: in statu nascendi). Thereupon, every time a certain rule of the grammar is applied, rewriting a non-terminal symbol A as a sequence of symbols $B_1B_2...B_k$ (with $k \geq 1$), this may be regarded as

* Cf. McCawley (1968).

an instruction for the addition of a subtree of the form:

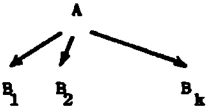


Figure 2.2

to the corresponding A-labeled node in the tree i.s.n..

Likewise in the case of the D-structure, as the first step, the initial symbol V is represented as the root of a D-tree i.s.n., and thereupon subtrees of the form

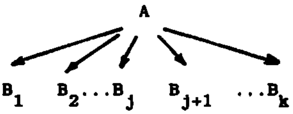


Figure 2.3

are added to the tree i.s.n. through application of D-rules of the form $A(B_1B_2 \ldots B_j * B_{j+1} \ldots B_k)$ with $k \geqslant 0$.

Thereby, in both cases, k new nodes, labeled B_1 to B_k are added to the tree i.s.n.. The relation holding between the A-node and the newly inserted B-nodes is called "direct dominance". Furthermore, the relation "is to the left of" is extended to all pairs (B_i,B_j) with $i < j$ and, for D-grammars moreover, to all pairs (B_i,A) with $1 \leqslant i \leqslant j$ and all pairs $(A,B_{i'})$ with $j+1 \leqslant i' \leqslant k$. This relation is reflected graphically in the left-to-right arrangement of the corresponding nodes.

It can be easily verified that the structures given in Figure 2.1, as far as they are connected by the "directly dominates"-relation, can be obtained by following these instructions along with the application of the well known set of rules $S \rightarrow NP + VP$, $VP \rightarrow V + NP$, $NP \rightarrow A + N$, $N \rightarrow$ man, ball..., $V \rightarrow$ hit..., $A \rightarrow$ a,the, in the case of the above C-structure and V(N*N), N(A*), A(*), in the case of the D-structure.

The above specification of the consequences of the "grafting operation" is only in terms of the internal structural properties of the grafted subtree. But there are external consequences as well, i.e. consequences for the relationships between the nodes of the added subtree and those of the tree i.s.n. to which it is added. These consequences can be expressed in the following way. Any node B in the added subtree relates to all nodes C outside the subtree in the same way as A does with respect to "is to the left of" and "dominates"*. That is, if A is to the left of C, then B is also to the left of C, if C dominates A, then C also dominates B.

In order to illustrate these "external specifications" let us assume that the formation process of the D-structure in Figure 2.1 (b) has got so far that the tree i.s.n. is

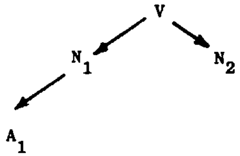


Figure 2.4

We now apply the rule N(A*) to N_2 and thereby attach the implied subtree

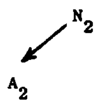


Figure 2.5

Since A_2 enters into the same external relations as N_2, the following implications hold: V dominates N_2 $\Rightarrow$ V dominates A_2; V is to the left of N_2 $\Rightarrow$ V is to the left of A_2; N_1 is to the left of N_2 $\Rightarrow$ N_1 is to the left of A_2; A_1 is to the left of N_2 $\Rightarrow$ A_1 is to the left of A_2.

So far, the great overlap in the characterization of C-structures and D-structures suggests a greater ressemblance between the two than in fact exists. There are essential differences between the two both in formal respect and in terms of the theoretical interpretation of various aspects of the formal structure. These differences are immediately apparent if one considers the two types of rules under concern. In the case of the above C-grammar the rules $A \rightarrow B_1B_2...B_k$ either specify a constituent's immediate subconstituents and their serial order in the replacing string, or they are used for paradigmatic specifications of a single category, as for instance subcategorization or lexical insertion with k=1 (e.g. V $\rightarrow$ hit, A $\rightarrow$ the). This is of course different in mixed, transformational models,

* A node X *dominates* a node Y if there is a sequence of nodes $Z_0Z_1,...,Z_n$ with $n \geqslant 1$ and $X=Z_0$ and $Y=Z_n$, such that Z_{i-1} directly dominates Z_i, with $i=1,...,n$.

which we, however, can ignore for the present discussion.

In the case of the D-grammar the rules $A(B_1B_2...B_j*B_{j+1}...B_k)$ specify that the lexical category outside the brackets, A, "is the head of" or "governs" the lexical categories B_i inside the brackets and, equivalently, that these latter categories "are the dependents of" or "depend on" the first. Hence, V(N*N) indicates that the verb category governs two nouns; the noun is specified as the head of the article by N(A*) and the article is specified as having no dependents by the rule A(*). Furthermore, the D-rule specifies the serial position of the dependents and also marks the position of the head category among its dependents by means of the asterisk. Unlike the C-rules, D-rules do not insert lexical entries. Lexical insertion is left over to another type of rules, specifying for every lexical category what terminal elements can be assigned to it. These conceptual differences between C-rules and D-rules reflect themselves in corresponding differences between C-structures and D-structures.

(1) The *nodes* in the C-structure are labeled by symbols from either the non-terminal vocabulary (S,NP,A, etc.) or from the terminal vocabulary (the, man,...). The nodes in the D-structure, as far as they are introduced by the grafting operations associated with the D-rules, are all labeled with *non-terminal* categories. These, moreover, are restricted to the preterminal, lexical type, for, unlike the C-grammar, the D-grammar contains no constituent symbols.

(2) In the C-tree the *branches* express two rather heterogeneous linguistic notions: (a) "has as one of its immediate constituents", (b) "is inserted by". In the D-tree this is not the case. D-tree branches all express the same linguistic notion "is the head of" or "governs" and not, like the branches in the C-structure, the relation "is inserted by". For lexical insertion another graphical means is used: the *dashed lines* that connect the lexical categories with the terminal elements assigned to them. There, of course, the "is the head of" relation is not meant to hold. These dashed lines therefore constitute an extra formal relation in the D-structure which is not independently specified in a C-structure.

(3) The C- and D-structures also differ in the completeness of the "is to the left of" relation specified between the nodes. In the D-structure this relation holds between all pairs of nodes. This is not so in C-structures. In a C-structure, the "is to the left of" relation is a partial order over the nodes, which is complementary with regard to the dominance relation.

Hence, there are pairs of nodes (A,B) such that neither A is to the left of B, nor B is to the left of A; in all these cases either A dominates B or B dominates A.

In order to anticipate the presentation and discussion of the interpretation theories to be dealt with, this section will conclude with a few indications of the relative importance of the diverse formal aspects of the structures to these interpretation theories. The most important formal aspect for the interpretation theories is the dominance relation. The predictions with regard to cohesion judgments will, both for C- and D-models, be based on formal notions which are defined in terms of the dominance relation.

As for the *nodes*, the *labels,* and the way the nodes are mapped onto the labels, these are of course important formal aspects as far as they indicate the localization, within the structure, of the elements which are mentioned in the predictions of cohesion judgments. In Levelt's interpretation theories these predictions are restricted to the terminal elements. On the other hand, nonterminal labeling and the information it yields concerning types of categories or constituents, plays only a marginal role in the interpretation theories, though it is sometimes referred to in the discussion.

Finally, the *is to the left of* relation, which is in fact more related to a grammar's weak than it is to its strong generative task, does not play a role in Levelt's interpretation theories. In the discussion thereof, however, it will be referred to, so it cannot be dispensed with entirely. In the last chapters of this study, moreover, signs will accumulate that a fully adequate cohesion model eventually will have to take word order into account.

2.2 COHESION JUDGMENTS; THEIR PLACE IN LINGUISTICS

In Chapter 1 we adopted Joan Bresnan's view of linguistic inquiry as an enterprise facing the two problems which she calls the characterization and realization problems. From this point of view the C- and D-structures can be regarded as two competing proposals for the *characterization* of syntactic structures. The question of whether we can make an empirically justified choice between the two, should be viewed as a matter of their relative *realizability* in models for performance data.

Furthermore, in Chapter 1 we became acquainted with cohesion judgments as those made by native speakers on whether or not, or to what degree, words or phrases belong together in a sentence. From a formal point of view, cohesion judgments can be compared to the well known similarity judgments in psychology. There, similarity judgments are fruitfully employed as manifestations of an individual's cognitive organization of the entities under concern. A large body of methods for the analysis of this kind of data has been developed by mathematical psychologists (see, for instance, Coombs, 1964; Torgerson, 1958; Runkel & McGrath, 1972, esp. Chapter 12 and the plethora of references given there; furthermore, in Dutch: Meerling, 1981, part 2). Much of this work also seems to be applicable to cohesion judgments, thanks to their formal correspondence to similarity judgments.

As a consequence, we feel confident -and expectant- in attributing to cohesion judgments an important role in the analysis of syntactic structure, and especially in resolving the above-mentioned choice between the two competing formalisms. The evaluation of the structural adequacy of either of the two formalisms is a problem exhibiting many aspects. One of these, which will be studied here, is the realizability of the competing formalisms in models for cohesion judgments.

In general, cohesion judgments do not play a prominent role in syntactic argumentation. Firstly, the divergent methodology (deviating from CTM) required by this type of observation is still unfamiliar among many linguists. Unfortunately, Levelt's far reaching and detailed proposals for a methodological approach to cohesion (see Section 2.3) have failed to change this situation. A notable exception, albeit along lines differing from those mapped out by Levelt, can be found in an impressive paper by Fodor, Garret, Walker and Parkes (1980), to which we shall have to return soon. Secondly, the lack of popularity seems to be related to the imagined invalidity of cohesion judgments. In proposing cohesion judgments as a standard for the evaluation of the structural adequacy of a given grammar, objections occasionally arise as to the presumed non-syntactic determinants of relatedness judgments. Syntactic structures, it is argued, will in all probability play a role in the process of forming cohesion judgments, but other determinants such as semantic similarity, surface distance between words, word associations etc. cannot be excluded.

Our general answer to this kind of objection has already been given in Chapter 1. Non-syntactic determinants of cohesion -in as far as they cannot be eliminated by careful data gathering procedures and experimental controls- might be explicitly accounted for in an integrated performance theory. In this case the problem of (in)validity assumes a different character since it becomes an inherent aspect of the construction of such a theory. The usefulness of cohesion judgments, in other words, does not necessarily depend on whether they are "perfectly valid" reflections of syntactic structure.

Nevertheless there ought to be indications of intimate relations between cohesion and syntactic structure. In this respect things promise well, as a review of some of the literature will show. We shall begin by taking a look at the above-mentioned study by Fodor et.al. (op.cit.), in which the question of the validity of cohesion judgments for syntactic distinctions is *explicitly* raised and receives an affirmative answer. Thereupon, we shall turn to some quotations in which linguists and psycholinguists refer to cohesion intuitions. These quotations strongly suggest that the analysis of cohesion judgments may help to make explicit considerations implicitly employed by linguists in deciding on structural descriptions.

(i) Fodor et.al. (op.cit.) argue against so called definitional decomposition of causatives in the underlying structure. In other words, they object to the derivation of surface constructions such as *John kills Mary* from underlying structures like *John causes Mary to die*. A crucial aspect of the putative definitional decomposition is that the surface arguments of the causative (here: *John* and *Mary*) are "homoclausal" in the surface structure, but "heteroclausal" in the underlying structure. In the authors' non-definitional account, however, *kill* is present in both surface and underlying structure, hence, *Mary* and *John* are homoclausal in both structures. The authors presented subjects with sentence pairs contrasting the causative constructions with non-causative counterparts, e.g. *John killed Mary* vs. *John bit Mary*. Cohesion judgments collected for critical pairs such as (John, Mary) and ($\left\{\begin{matrix}\text{killed}\\ \text{bit}\end{matrix}\right\}$, Mary) did not vary significantly over the sentences. The authors felt justified in regarding this as counterevidence against the definitional account *in virtue of a prior validation* of the test procedure. In the validation phase of the experiment the investigators had presented the subjects with sentence pairs of various linguistically non-problematic types like *I expected John to leave* and *I persuaded John to leave*. All these pairs exhibited the contrast between underlying homo- and heteroclausality, but none of them were of the putative definitional type. Systematic variation over the members of the pairs indicated cohesion judgments to be sensitive to the crucial contrast.

(ii) Various passages in the literature encourage confidence in the relevance of cohesion judgments. Both linguists and psycholinguists have repeatedly offered their opinions on sentence structure together with introspective observations on the degree of relatedness, cohesiveness, connectedness, tiedness and so on, of the words or phrases comprising a sentence. Chomsky's passage, referred to on the first page of this study is one example, others are contained in the following review of quotations.

Hays (1964, p. 525) discusses various aspects which in his opinion are all relevant to the problem of justifying a choice between C- and D-theory. In this discussion he writes:

> Another line of interpretation makes dependency a psychological relation. Empirically, this interpretation would require data about the structural intuitions of native speakers; dependency theory would be supported if many or most native speakers could agree on the central (governing) element in each utterance presented to them, and on the connections binding elements together. Phrase-structure theory would be supported if they agreed on containments, e.g. THAT OBJECT-VERB RELATIONS ARE CLOSER THAN SUBJECT-VERB RELATIONS, so that predicates are contained in sentences*.

A second example, with an extension of the notion of relatedness strength over morphemes, is taken from a study by Jane Robinson (1969, p. 63). She disagrees with Fillmore's (1968) partitioning of the sentence into a proposition constituent, a tense-less set of case relations between the nouns and the verb, and a modality constituent, accounting for tense, mode, negation and aspect. According to her, negation and aspect are to be included in the proposition constituent as properties of the verb, for ...

> Briefly, one reason is that negation and aspect ARE MORE CLOSELY TIED TO THE VERB and accompany it under nominalizing transformations while tense, mode, and interrogation morphemes do not*.

G. Miller (1962, p. 749) introduces the notion of syntactic structure with the following words:

> Take the sentence *Bill hit the ball*. To native speakers of English it is intuitively obvious that this sequence of words has a kind of structure, that some pairs of adjacent words ARE MORE CLOSELY RELATED THAN OTHERS. For instance *the ball* feels like a more natural unit than others*.

Lyons (1969, p. 202 ff.) uses the term "cohesion" several times in his discussion of the "word" and its status in general linguistics. He subordinates the justification and definition of the word as a unit intermediate in rank between the morpheme and the sentence to the primary purpose of grammatical description: to generate the sentences of a language. Part of this task in-

* Capitals mine, E.S.

volves the combination of several morphemes into units forming grammatical words, a procedure which acknowledges the fact that within words "these morphemes ARE IN GREATER 'COHESION' than other groupings of morphemes in the sentence which are not recognized as words".* According to Lyons, other attempts to define words in terms of semantic properties, potential pauses or occurrence as a minimal free form (Bloomfield's criterion) are all based on secondary correlations of the central notion of structural cohesion. So, for instance, argues Lyons:

> ... as a matter of empirical fact it may be true that the set of 'minimal free forms' will generally correspond in all languages to the set of phonological units representing grammatical words; but if so, this fact presumably depends upon and reflects THE STRUCTURAL 'COHESION' OF THE WORD IN SENTENCES, and is of only indirect concern to the grammarian*.

The term "cohesion" even appears in the title of a book by M.A.K. Halliday and Ruqaiya Hasan (1975): "Cohesion in English". Unlike our own use of the term, "cohesion" is there employed as a theoretical expression for semantically based relations which distinguish a text from a set of unrelated sentences. In general, though not exclusively, it refers to relationships holding across sentence boundaries. This restriction of "cohesion" to the semantic determinants of texture is accompanied by the acknowledgement (see p. 7) that from a more general point of view, the phenomenon can be more broadly conceived:

> In general, any unit which is structured hangs together so as to form text. All grammatical units - sentences, clauses, groups, words - ARE INTERNALLY 'COHESIVE' simply because they are structured*.

We complete this review with a striking quotation from Tesnière (1953, Chapter 16, verses 7 to 11):

> 7.-Il résulte de ce qui précède que la morphologie est essentiellement et uniquement l'étude des marquants.
>
> 8.-Les marquants diffèrent entre eux par trois caractères: leur nature, leur ordre, leur adhérence.
>
> 9.-La nature des marquants est le vêtement phonétique qui les constitue.
>
> 10.-L'ordre des marquants est celui dans lequel ils se succèdent sur la chaine parlée. Il n'est donc que la réplique morphologique de

* Capitals mine, E.S.

l'ordre linéaire (...).

11.-Enfin l'adhérence des marquants est LE DEGRE DE COHESION qui unit entre eux ceux qui sont en séquence sur la chaîne parlée. L'adhérence est donc fonction inverse de la profondeur des coupures (...)*.

Earlier in Tesnière's book, the depths of the word boundaries (la profondeur des coupures), inversely related to the degree of cohesion between words, are claimed to correspond to what the author refers to as "la hiérarchie des connexions". In effect, the structural descriptions that Tesnière assigns to sentences are essentially D-trees in which the "hiérarchie des connexions" is reflected by the "height" of the branches. Further details of this correspondence can be found in the original text and need not concern us here. It would, however, detract from Tesnière's merits if we failed to recognize the above passage as a prototype of an interpretation theory -avant la lettre- for cohesion in terms of a D-structure.

In concluding this section, we should bear in mind that when cohesion judgments are used for the evaluation of grammars, a distinction should be made between their application in either a *broad* or a *narrow* sense. This becomes clear when we compare the above-mentioned examples from Chomsky (1965) and Hays (1964) respectively. In Chomsky's example, the cohesion observation plays a role in an option between possibilities existing *within* one and the same formalism, i.e. C-theory. In Hays' example, it is suggested that they can also be used for deciding options *between* formalisms. Both applications will be illustrated further in the section on interpretation theories.

2.2.1 Cohesion judgments; methods of collecting them

The broad characterization of cohesion judgments as those concerning the degree of relatedness between the words or phrases comprising a sentence, can be further specified in terms of the procedures used to collect them. The interpretation theories to be dealt with in this and most other chapters in this book, predict a rank order over some or all of the word pairs (or pairs of phrases) with respect to cohesiveness. We therefore require cohesion observations yielding an empirical order relation over word pairs with regard to perceived cohesiveness. This requirement renders cohesion data formally com-

* Capitals mine, E.S.

parable with the well known similarities data in psychology, which Coombs (1964) defines as order relations on pairs of dyads whose objects are all from the same set. An agreeable consequence of this formal correspondence is that much of the experience acquired by mathematical psychologists in collecting and analyzing similarities data applies to cohesion data as well. This would seem to hold for all data *collecting* methods, though some reservation should be made about methods for *analyzing* the data. It is, for instance, on the ground of the conceptual differences between similarity and cohesion, doubtful whether multidimensional scaling methods could be of any use for the analysis of cohesion. Cohesion in a sentence -unlike similarity- can hardly be grasped in terms of dimensionality.

It is, however, data gathering rather than data analysis which concerns us here. Further considerations on data analysis will therefore be postponed until we come to deal with the interpretation theories with which it is intimately involved. As already mentioned, we require cohesion observations which yield data that can be formally equated with the kind Coombs refers to as "similarities data". As this phrasing intentionally suggests, we shall make a distinction between the primary observations made and the data to be analyzed. We do not observe an order relation in itself, but we shall make our observations so as to allow for an interpretation in terms of an order relation. In making this distinction we follow Coombs (1953, p. 470):

> What one 'finds out' from one's data is a function of two things: the information in the data and how this information is extracted. What information the data contain depends on how it is collected. Some methods of collecting data 'permit' more characteristics of behavior to exhibit themselves than do other methods. Or, in opposite terms, some methods of collecting data impose properties on the behavior that other methods do not."

In this vein, we must recognize that the empirical order we wish to analyze is an order at the phenotypic level (see Chapter 1). In choosing a particular method of data collection, the investigator intrusively restricts the amount or the nature of genotypic information capable of manifesting itself phenotypically. This choice of method, determined in part by theoretical objectives, in part by practical considerations, will therefore impose constraints on the possible inferences that can ultimately be drawn from the analysis.

From this perspective let us take a look at the three methods of collecting cohesion judgments described by Levelt (1974c) as the most common in the literature: (i) rank ordering of word pairs, (ii) assigning scale values to word pairs and (iii) word sorting.

Rank ordering of word pairs. In the first method, the subject is presented with every dyad of word pairs (i, j) vs. (k, l) of a sentence, with i, j, k, l ranging over the sentence's word occurrences, and asked to indicate which word pair coheres more strongly. When the subject chooses (i, j), we write (i, j) $\gtrsim$ (k, l). This relation should be read "i and j cohere at least as strongly as k and l."

The set of word occurrences together with the ordered dyads of word pairs ((i, j), ((k,l)) for which this relation holds, constitute the empirical relational system we wish to analyze Ordering word pairs is both the most direct and most laborious way of obtaining the desired information. For a sentence consisting of n words there are $\binom{n}{2} = n(n-1)/2$ word pairs and $\binom{\binom{n}{2}}{2} =$ $= n(n-1)(n^2-n-2)/8$ pairwise comparisons of word pairs (henceforth PWCs). The complete method of pair comparisons would, for instance, require 45 PWCs for a five word sentence and 990 PWCs for a ten word sentence. Hence, even were there reasons for preferring the method of pair comparisons, we would soon reach the point where practical considerations force us either to choose labour saving variants, such as the method of triadic comparisons (see Levelt, op.cit.), or to switch to alternative methods altogether.

Assigning scale values to word pairs. As an example of the second method, viz. assigning scale values to word pairs, we could have subjects rate the word pairs on a *seven point scale* according to intuited relatedness. In this situation we might decide to infer the empirical relation, $\gtrsim$, from the scale values, s(i, j), assigned to the word pairs (i, j) in the following way: for all words i, j, k, l, s(i, j) $\geq$ s(k, l) $\Leftrightarrow$ (i, j) $\gtrsim$ (k, l). It is clear that we shall arrive at the relation $\gtrsim$ with much less effort. The reduction in number of presentations from the method of pair comparisons to the method of rating scales goes from 45 to 10 for a five word sentence and from 990 to 45 for a ten word sentence.

As already pointed out, however, it is not only practical considerations which determine the choice among data collecting methods. Theoretical considerations are involved as well. The above two procedures for arriving at the empirical relation, $\gtrsim$, differ in the way they permit properties of the data to exhibit themselves. Here we shall not attempt completeness, but restrict ourselves to mentioning the issues of *transitivity* and *interdependence* of choices. In the seven point scale procedure, since the empirical relation $\gtrsim$ takes over the ordinal properties of the relation $\geq$ over the assigned scale values (i.e. reflexivity, antisymmetry and transitivity), the relation is of necessity transitive. Thus for all triples of pairs (i, j), (k, l), (m, n): (i, j) $\gtrsim_{s}$ (k, l) and (k, l) $\gtrsim_{s}$ (m, n) $\rightarrow$ (i, j) $\gtrsim_{s}$ (m, n). This is not necessarily the case for the relation $\gtrsim_{s}$ when determined by the method of pair comparisons. Hence, any underlying intransitivity in the choice process of the subject, due -for instance- to fluctuation of judging criteria, can only become manifest in the method of pair comparisons and not in the seven point scaling method. Conversely, the relation $\gtrsim_{s}$, as obtained by means of pair comparisons may violate a predicted order by not being an order at all. In the case of rating scales, however, the possibility of violation is restricted to "being a deviating order". In making a choice between the above procedures, the investigator must choose, as Coombs (1953, p. 486) elsewhere says: "... between *mapping* his data into a simple order and *asking* his data whether they satisfy a simple order."

A related theoretical consideration is the difference in the degree of interdependence between the $\binom{\binom{n}{2}}{2}$ instances of the empirical relation $\gtrsim_{s}$. In the case of the seven point scale procedure these instances are implied by the $\binom{n}{2}$ assigned scale values and thus based on $\binom{n}{2}$ separate decisions. The $\binom{\binom{n}{2}}{2}$ PWCs induced by the seven point scale procedure are therefore much more interdependent than those resulting from the method of pair comparisons, where the number of separate decisions is $\binom{\binom{n}{2}}{2}$.

In Chapter 5, probabilistic interpretation theories will be presented specifying for every PWC (i, j) versus (k, l) the choice probability p((i, j) $\gtrsim_{s}$ (k, l)). Since the statistical procedures proposed for testing these theories assume independence of the choices over PWCs, the seven point scale procedure would be inadequate for obtaining the empirical choice relation.

Word sorting. The constraints, preventing phenotypical manifestation of genotypical characteristics underlying behaviour, increase even more when we turn from seven point scales to the method of *word sorting*. In this method the subject is presented with a sentence and with cards on which the separate words have been written. The subject's task is to partition the cards into mutually exclusive subsets according to the intuited cohesiveness. Words that cohere should end up in the same pile, words that do not cohere should, as far as possible, find themselves distributed over different piles. The result of a single subject's word sorting can be recorded in an "adjacency matrix", with rows and columns corresponding to the words, and cells, corresponding to the word pairs. The cells of the matrix contain values s(i, j), equal either to 1 or 0, depending on whether or not the words i and j are present in the same pile. Generally, the empirical choice relation $\gtrsim$, over the word pairs is obtained by a process of first summing all adjacency matrices, thus yielding a matrix whose cell-values express the numbers of subjects who put i and j into one pile, and then determining the order over these numbers.

Let us first consider the properties of a single subject's matrix. We use the sentence *the girls left* as an example and assume that a given subject's partitioning is: {the, girls} and {left}. The corresponding s-values are s(the, girls) = 1, s(the, left) = s(girls, left) = 0, and the derived empirical choices are (the, girls) $\succ$ (the, left) $=_o$ (girls, left).*
The most striking difference between the methods of word sorting and the two afore-mentioned procedures for arriving at the order $\gtrsim$, comes to light when we summarize the possibilities for any triple of words (i, j, k) in a single subject's partitioning. These are: (i) i, j and k are put into the same subset, (ii) i, j and k are put into three different subsets, (iii) two of the words, say i and j are put into the same subset with the third in another. This implies that for all triples of words *at least two of the s-values are equal to each other and the third is either equal to (cases i and ii) or greater (case iii) than these two values*. This means that the s-values (and the derived choice relation) necessarily meet the so called ultrametric inequality (see Levelt, 1974c, p. 41; Johnson, 1967), requiring for all triples of words i, j, k that s(i, j) $\gtrsim$ min [s(i, k); s(j, k)]. The consequence of this is, as we will discuss in the section on Levelt's interpretation theories, that a single subject's choice relation necessarily satisfies the necessary and sufficient condition for representability by some C-structure (at least as far as Levelt's interpretation theory is concerned). This, of course is not indicative of genotypic "constituency".

On the other hand, the word sorting procedure masks characteristics of behaviour at the "genotypic" level, that might ask for a D-structure approach. Experience with cohesion judgments has taught us that the relation between a word and the *head* of an *endocentric* construction are intuited as more cohesive

*) Here $\gtrsim$ is interpreted as a reflexive, antisymmetric (see below) and transitive relation of weak dominance. Hence it can be decomposed into the irreflexive, asymmetric and transitive relation of strict dominance, $\succ$, and the reflexive, symmetric and transitive relation of equivalence, $=_o$. (i, j) $\succ$ (k, l) holds whenever (i, j) $\gtrsim$ (k.l) and not (k, l) $\gtrsim$ (i, j). Furthermore, (i, j) $=_o$ (k, l) holds, by antisymmetry, whenever (i, j) $\gtrsim$ (k, l) and (k, l) $\gtrsim$ (i, j).

than the relation between that word and the *modifier* of such a construction. Hence, if a subject's genotypic order over the word pairs of *the girls left* is, from strong to weak: (the, girls), (girls, left), (the, left), then both the method of pair comparisons and the method of seven point scales permit this order to become manifest at the phenotypic level. But it is impossible to express the strong cohesion of (the, girls), the intermediate cohesion of (girls, left) and the weak cohesion of (the, left) simultaneously in a single partitioning of the word sorting method.

Thus far, our comparison of the procedures for arriving at an empirical choice relation over word pairs has been restricted to the level of a single subject's task. The picture will become modified however, when, as is usually the case in cohesion research, the empirical choice relation is inferred from the choices of a *group* of subjects by means of a collective choice rule, such as a majority decision. If, for instance, two of three subjects were to partition *the girls left* in {the, girls} and {left}, and one of them in {the} and {girls, left}, the sum of the corresponding adjacency matrices would be:

	the	girls	left
the	3	2	0
girls	2	3	1
left	0	1	3

The majority rule would yield the collective ordering, from strongly to weakly cohesive: (the, girls), (girls, left), (the, left). This no longer satisfies the ultrametric property of the individual orderings.

Furthermore, as regards the seven point scale procedure, the "paradox of voting" teaches us that application of the majority decision to individual (transitive !) orderings may yield an intransitive collective choice relation. Assume, for instance, that three subjects A, B and C assign to the word pairs of the sentence *Mary is happy* the ratings given in the following matrix:

	(Mary, is)	(Mary, happy)	(is, happy)
A	5	3	7
B	4	6	5
C	7	6	5

Here, the majority decisions for the PWCs of these pairs: (Mary, is) vs. (Mary, happy), (Mary, happy) vs. (is, happy) and (is, happy) vs. (Mary, is) are all two to one, yielding an intransitive collective choice relation.

Choice of method. Further aspects of the data collecting methods need not concern us here; they will be dealt with later when needed. The above comparison of methods has been restricted to those aspects which we have to consider in selecting the one most suited to the purpose at hand. The various interpretation theories are formulated so as to predict for every PWC of word pairs or pairs of phrases (either deterministically or probabilistically) the choice that will be made. In general, we want to be able to test these predictions as independently as possible. This aim makes the method of pair comparisons the most attractive of the three. To the extent that interpretation theories are probabilistic, the methods for data analysis and the statistical testing procedures even explicitly require independence of the choice probabilities over PWCs. Whenever it is feasible, therefore, this is the method we shall use.

There will, however, be instances where practical considerations force us to abandon the method of pair comparisons. In Chapter 4 an explorative empirical comparison will be made of several deterministic interpretation theories predicting the choices for PWCs in which both word pairs and constituent pairs are involved. Some of the analyses would require more than 250 presentations to a subject if the complete method of pair comparisons were to be applied. In these cases seven point scales will be used, together with a collective choice rule to be described later.

Finally, the word sorting method has been seen to endow behaviour with properties which are crucial for representability by means of C-structures and masks properties that might favour a D-structure approach. The method is therefore inappropriate to our main purpose: comparison of the structural adequacies of C-theory and D-theory. Although we have seen that collective choice relations obtained by word sorting can be non-ultrametric, experience has shown that it is strongly biased towards being ultrametric. This is because the individual orderings that constitute its building blocks are in themselves all ultrametric of necessity. Thus, the more similar the behaviour of the subjects, the more the group result approaches ultrametricity, irrespective of whether or not the individuals' genotypic behaviour is ultrametric.

2.3 LEVELT'S INTERPRETATION THEORIES

Both Levelt's constituent model for cohesion judgments, henceforth denoted as CGL, and his dependency model, DGL, consist of two components: a linguistic theory and an interpretation theory for relating the former to cohesion judgments. It goes without saying that the linguistic components of CGL and DGL are C- and D-grammars, respectively. The interpretation components of both models can be viewed as, firstly, defining a distance metric over the structures, and secondly, as deriving in an inverse fashion the cohesiveness order over the word pairs from the order over the distances between the word pairs. Superficially seen, the first part of this characterization seems to apply only to DGL and not to CGL. In the case of DGL, the author himself has explicitly pointed out in a footnote (Levelt, 1974c, p. 53) that the interpretation theory adds a distance model to the linguistic theory. Strictly speaking, as it has been formalized, CGL does not contain a distance model. However, by introducing insignificant modifications which still preserve the underlying idea, it is possible to reformulate it as a distance model in which the same conditions for representability by C-structures are demanded from the data (see the concluding paragraph of Section 2.3.1). In the two sections to follow we shal deal with the interpretation theories in some detail. After the essentials of CGL have been described, the empirical reasons which brought the author to modify and ultimately abandon the model will be discussed. Following this we shall turn to DGL, seemingly more promising in that respect where CGL shows its main shortcomings. As far as possible our own criticism will be postponed until the last section of this chapter.

2.3.1 Levelt's constituent model for cohesion judgments

In his section "A constituent model for relatedness judgments" Levelt attempts two versions of CGL. In one of these the linguistic component is a C-grammar in the form of a pure, i.e. exclusively nontransformational model. In the other version, the linguistic component is a transformational grammar with a C-structure base, as in Chomsky's Aspects. Both versions contain the same interpretation component. This has already been referred to informally in the discussion of Chomsky's cohesion observation given at the very beginning of this study. There the interpretation was said to relate the cohesion between words, in an inverse fashion, to the height of their first dominating node. This basic idea is formalized in CGL in three steps.

The first step is the definition of a real-valued COHESION-FUNCTION α over the nodes of a C-structure, such that whenever node A dominates node B, then $\alpha(A) < \alpha(B)$, for all nodes A, B in the C-structure. The cohesion of a constituent X, $\alpha(X)$, is equated with $\alpha(K)$, where K is the node that represents X. As a consequence, along every path leading from the root to the terminal nodes, the cohesion values assigned to the nodes are strictly increasing, and all constituents are necessarily less cohesive than their subconstituents.

As a second step, the notion is introduced of the SMALLEST COMMON CONSTITU-

ENT of a word pair (i, j), to be denoted as SCC(i, j), as that constituent represented by the lowest node dominating both i and j.

The third step is the formulation of an *interpretation axiom* expressing the relatedness r(i, j) between the words i and j in terms of the cohesion of the SCC(i, j). The interpretation axiom specifies the cohesiveness order over the word pairs by postulating a monotone relationship between the r-values and the α's of the relevant smallest common constituents. This axiom reads as follows: "For all words i, j, k, l in the sentence, $r(i, j) < r(k, l) \Leftrightarrow \alpha(SCC(i, j)) < \alpha(SCC(k, l))$". The intention thereby is that equal degrees of relatedness will correspond with equal cohesion values and vice versa. There is, however, some indeterminacy as to whether this really follows from the axiom by exclusion, as stated by the author. This is due to the fact that the cohesion function α is not unambiguously defined by the dominance relation. When the smallest common constituents of two word pairs (i, j) and (k, l) appear in the "is to the left of" relation, no restrictions on the α's result, since this relation and the dominance relation are complementary. An alternative formulation, therefore, might directly derive the incomplete relations of cohesiveness order and cohesiveness equivalence over the word pairs. This might be implied from the partial dominance order over the relevant SCCs without the intermediate introduction of the indeterminate cohesion function α: SCC(i, j) dominates SCC(k, l) $\Rightarrow$ $r(i, j) < r(k, l)$; SCC(i, j) = SCC(k, l) $\Rightarrow$ $r(i, j) = r(k, l)$.

Levelt illustrates his interpretation theory by the following example:

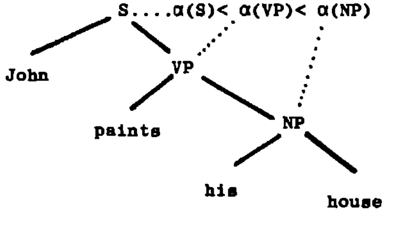


Figure 2.6 C-structure for the sentence *John paints his house* with cohesion values assigned to the nodes

The relatedness between *John* and *house* r(John, house) is associated with the cohesion of the SCC of *John* and *house*, viz. S; the relatedness between *paints* and *house* with the cohesion of the SCC of *paints* and *house*, viz. VP. Since, according to the definition of cohesion $\alpha(S) < \alpha(VP)$, application of the interpretation axiom for this PWC results in the prediction: r(John, house) < r(paints, house). Table 2.1 summarizes the predictions for all the pair-wise comparisons of word pairs in this sentences. The table gives the SCC's for all word pairs and further specifies whether the row word pair is less related than (<) or equally related to (=) the column word pair.

Table 2.1 Matrix of predictions of CGL, applied to the structure of Figure 2.6

	pair	j,p	j,hi	j,ho	p,hi	p,ho	hi,ho
Pair	SCC	S	S	S	VP	VP	NP
John, paints	S	-	=	=	<	<	<
John, his	S		-	=	<	<	<
John, house	S			-	<	<	<
paints, his	VP				-	=	<
paints, house	VP					-	<
his, house	NP						-

In this way, from any proposed C-structure, application of CGL will derive a set of predictions that can be tested against the empirical results of a judging experiment. Given the interpretation theory a numerical index such as the percentage of violations could serve for an evaluation of the goodness of fit of the proposed C-structure. Pages 37 and 38 of Levelt's (1974c) book will provide the interested reader with an impression of how such an evaluation of alternative C-structures could proceed.

Instead of asking which tree best fits the data for a given interpretation theory, we may pose the more fundamental question of whether, within the restrictions of the model, any tree exists fitting the data. This question could be answered trivially by evaluating all possible C-structures for a sentence of given length in the above manner. This laborious, brute force procedure, however, would not prove very informative as to the critical factors governing the data's representability or non-representability by means of a C-structure. It would, therefore, be preferable to try to identify the critical property necessary if a data matrix is to render a representation according to the model.

Levelt shows that in order to be representable by a C-structure according to the above interpretation theory, a data matrix should satisfy the so called

ULTRAMETRIC INEQUALITY (see Johnson, 1967). This means that for all triples of row (column) elements i, j, k of the data matrix, we should require that $r(i, k) \geqslant \min [r(i, j); r(j, k)]$. In the ideal errorless case, this necessary and sufficient condition requires for all triples of words i, j, k, that two of the relatedness values should be equal to each other, whereas the third must be greater than or equal to these two r-values. It is not difficult to imagine a computer algorithm checking every triple of elements to see whether this ultrametric inequality is met by the data. But it is clear to reflection that no matrix of empirical relatedness values can really be expected to meet this severe requirement in any strict sense. It goes without saying that a test for ultrametricity in a data matrix should be executed while taking statistical account of the ever present measurement error (for this problem, see Loosen, 1972).

Reformulation as a distance model. At this juncture it is informative to observe how a trivial modification could allow the interpretation axiom to predict values of unrelatedness u(i, j) instead of relatedness r(i, j). This insignificant modification could be introduced without consequences for the underlying basic idea. The axiom, then, would read like this:

$u(i, j) > u(k, l) \Leftrightarrow SCC(i, j)$ dominates $SCC(k, l)$

$u(i, j) = u(k, l) \Leftrightarrow SCC(i, j) = SCC(k, l)$

(for all words i, j, k, l)

The consequence for the formulation of the ultrametric inequality would be: $u(i, k) \leqslant \max [u(i, j); u(j, k)]$ (for all words i, j, k).

This would be recognizable as a special case of the triangular inequality, which states that $u(i, k) \leqslant u(i, j) + u(j, k)$. Moreover, since $u(i, j) = u(j, i)$ for all words i, j, we see that u satisfies two of the properties of a distance metrix (for this notice see Section 2.4.1). After the addition -as a technical assumption- of: $u(i, j) = 0 \Leftrightarrow i = j$, u qualifies as a distance metric.

2.3.1.1 Levelt's rejection of the constituent model

Levelt concludes his section on the C-model for relatedness judgments with a discussion of some empirical analyses performed in order to test the model. Two versions of CGL have been put to the test. In one version the linguistic component is exclusively comprised of a C-grammar. In the second version, introduced by the author in order to overcome certain shortcomings in the first version, the linguistic component is a transformational grammar taking a C-grammar as its base, as in Aspects. Accordingly, we shall refer to these versions as "pure CGL" and "mixed CGL" respectively.

As a result of the analyses, both "pure CGL" and "mixed CGL" are rejected as inadequate for the representation of cohesion judgments. "Pure CGL" could have been expected to fail on the ground of a general principle discovered by Levelt soon after the start of his cohesion research: relatedness judgments express underlying relations among the words of a sentence (see, for instance, Levelt, 1969, 1970). Furthermore, both "pure" and "mixed CGL" are affected by the systematic violation of the ultrametric inequality, even in cases where this can not be accounted for by the above principle. Both reasons for rejecting CGL will now be discussed in some detail.

In the handling of C-grammar as a *pure*, non-transformational model, the well known distinction between underlying and surface structure becomes ignored. One consequence of this is that the entities about which cohesion judgments are gathered, say words and phrases, directly correspond to terminal elements or constituents in the C-structure. No extra interpretation is required in order to connect structural elements with data elements. In this case, if the data were to systematically violate the ultrametricity, this would provide an immediate argument for the rejection of the pure C-model, given the maintaince of Levelt's interpretation. As a matter of fact, violations of various types abound in almost all judgment tasks.

Many such violations concern cohesion judgments collected for sentences to which the Aspects model would assign a deep structure differing strongly from the surface structure. Moreover, they strongly suggest that it is mainly this deep structure that determines the cohesion judgments. These types of violation stress the untenability of pure CGL. See, for instance, Levelt's (1974c, p. 42) experiments using sentences from which certain words have been deleted transformationally. On the basis of this counterevidence, Levelt abandoned "pure CGL" and decided to try the afore-mentioned "mixed CGL".

In this second model the interpretation axiom and the cohesion function remain unchanged; these, however, no longer operate on the surface C-structure, but on the deep C-structure, whose relations are now considered to determine the cohesion judgments. The investigation of the second model is complicated by absence of the one-one correspondence between structural elements and empirical entities constituting the pairs to be judged. The empirical entities are *words* and *phrases*, which can be said to more or less correspond with particular elements or constituents of the surface structure.

The model, however, bases its predictions on a sentence's deep structure, whose elements and constituents are not in one-one relation with those of the surface structure. This means that extra "translation work" must often be performed before the putative agreement between deep structure relations and cohesion judgments can be established. A couple of examples may clarify this.

There are, for instance, transformational fusions between elements that are separated in the deep structure. In a sentence like *he reads no book* the surface element *no* corresponds to two deep structure elements: the negation formative, say *not*, and the numeral *one* (or, in an alternative version, the artical *a*). How are word pairs including *no* as one of their members to be accounted for by the deep structure relations as formulated in "mixed CGL"? At the least, some degree of extra interpretation is needed.

As a second example we can take a look at sentences from which elements of the underlying structure have been deleted transformationally. An example of this is the sentence *John eats apples and Peter pears** which has been transformationally derived from *John* $eats_1$ *apples and Peter* $eats_2$ *pears*. The second *eats* has been erased. This is one of the sentences in Levelt's critical experiments which led to the rejection of pure CGL and stimulated a "mixed CGL" approach. But before the data can be shown to correspond to the underlying C-structure, an ad hoc elaboration of the interpretation theory must be given in order to connect deep structure elements with the surface elements concerning which the data are gathered. This extra interpretation charges the surface *eats* with a double role: in some pairs it is considered as a reflection of $eats_1$ in the deep structure, in others as a reflection of $eats_2$. Following this extension of the interpretation theory, a fair agreement between deep structure relations and cohesion judgments can be established.

Such examples stress the necessity of formulating explicit rules and incorporating these into the interpretation theory before drawing up the balance sheet of the virtues and vices of "mixed CGL". Nevertheless, in spite of this indeterminacy regarding "mixed CGL", there is a clear reason for rejecting it.

This rejection cannot be simply based on showing that the data is not ultrametric. The absence of ultrametricity in data, collected over elements

* I apologize to the revisor of my English for not having followed his suggestion to have Benjamin Britten eat the apples and Peter pears : Prof. Levelt, whom I had to quote, had John to be the agent of the eating.

of a surface structure interpreted as a transformationally strongly modified deep structure, does not necessarily falsify ultrametricity in the deep structure. The ultrametricity, however, remains a vital condition to be imposed on cohesion data, concerning that class of sentences in which this transformational modification is minimal. We refer here to the class of simple declarative sentences for which a rather direct correspondence between deep and surface structure elements can be established, without the necessity of extra interpretation.

Now, the evidence shows that ultrametricity systematically fails to hold, even for these simple declarative sentences. Levelt has pointed out that it is especially the presence of endocentric constituents which causes severe violations of the predicted equalities. For judgments of PWCs (i, k) versus (j, k), such that i and j are the *head* and *modifier* of an endocentric construction respectively, and k a word outside this construction, it is stereotypically the case that r(i, k) greatly exceeds r(j, k). But in these cases the model predicts equalities. We can illustrate this by returning to the sentence *the girls left*, in which *the* and *girls* form the modifier and head of an endocentric constituent respectively, and *left* the element outside the constituent. The C-structure for this sentence is:

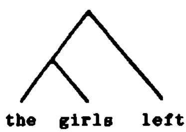


Figure 2.7

and the predicted empirical relations are:

r(the, girls) > <u>r(the, left) = r(girls, left)</u>.

The empirical choices, in all probability, will be:

r(the, girls) $\gg$ <u>r(girls, left) $\gg$ r(the, left)</u>.

The characteristic violation has been marked by the underlined PWC.

Following these considerations, we must conclude that CGL is inadequate as a model for relatedness judgments. It should be noted that the rejection of CGL is in fact a rejection of the conjunction of a particular linguistic theory, the C-grammar, with a particular interpretation theory. From a logical point of view this means that either the constituent model, or the interpretation theory, or both, are untenable. Levelt appears to be inclined to the first conclusion, although he explicitly states that the rejection of the C-grammar is conditional upon the maintainance of the interpretation theory.

Accordingly, he acknowledges the right of a linguist to "set these findings aside by rejecting the interpretation theory" (Levelt, 1974c, p. 63). On the other hand, he challenges this linguist to come up with an alternative interpretation for this case. It is precisely this challenge which forms the stimulus for much of this study. In anticipation of the last section of this chapter, this is a natural context to inform the reader that we do not subscribe to the above interpretation of the C-grammar. Consequently, in dismissing the argumentation given for the rejection of CGL, the need of an alternative interpretation theory makes itself felt. This alternative will be presented in Chapter 4.

For the moment, however, we shall continue to pursue Levelt's argument. The main reason for the rejection of CGL was its failure to account for the above indicated pairwise comparisons involving the head and modifier of an endocentric construction. According to Levelt this incapacity is not surprising. A C-grammar is not suited to express the asymmetric relation of dependency holding between head and modifier of an endocentric construction. And since this dependency relation is explicitly accounted for by D-grammar, the hope can be entertained, that ... "An obvious alternative is to use a dependency grammar as a linguistic theory, and to adapt the formulation of the interpretation axiom accordingly" (Levelt, op.cit. p. 50).

2.3.2 Levelt's dependency model for cohesion judgments

Since the evidence indicated that cohesion judgments are determined mainly by the underlying deep structures, Levelt decided against any kind of "pure DGL". He supposes that the linguistic component of DGL must be a transformational model, the base of which, of course, is a D-grammar. The interpretation component consists of several definitions and, again, an interpretation axiom, relating the formal model to the judgments.

Firstly, a real-valued DEPENDENCY function α is defined over the nodes of a D-structure, such that whenever node B directly depends on node A, then $\alpha(A) < \alpha(B)$, for all nodes A, B in the D-structure. Secondly, by convention, every element in the D-structure is declared to depend on itself. This paves the way for the introduction of the FIRST COMMON HEAD of a pair of nodes, A, B in the D-structure, symbolically FCH(A, B), as the lowest node on which both A and B depend. The next step is the introduction of the distance-like notion of disconnectedness: "The DEGREE OF DISCONNECTEDNESS of two elements A and B, $\delta(A, B)$ in a dependency diagram is defined as follows:

$$\delta(A, B) = [\alpha(A) - \alpha(FCH(A, B))] + [\alpha(B) - \alpha(FCH(A, B))] =$$
$$= \alpha(A) + \alpha(B) - 2\alpha(FCH(A, B))\text{"}$$

Finally, the interpretation axiom, specifying an inverse relation between cohesiveness and disconnectedness, reads:

"$r(i, j) < r(k, l) \Leftrightarrow \delta(i, j) > \delta(k, l)$

(for all words i, j, k, l in the sentence)"

Levelt illustrates his interpretation theory by reference to the D-structure for the sentence *the pianist plays beautifully*, which is given in Figure 2.8. In order to simplify this structure, we have, by convention, replaced the lexical categories in the D-structure by the words that have been assigned to these categories in the derivation of this sentence. These words are regarded as standing in the same relations to each other as those holding between the lexical categories they have replaced. Furthermore, sums of parameters, p, q, ..., all of them assumed to be positive, have been assigned to the nodes in such a way as to reflect the partial order of their dependency values α according to the above definition.

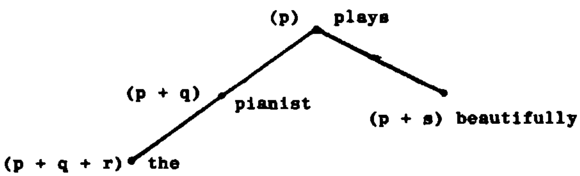


Figure 2.8 Simplified D-structure for the sentence *the pianist plays beautifully*, with parametric representation of dependency values

As an example we shall derive the prediction of the cohesiveness order for the PWC (pianist, plays) versus (pianist, beautifully) in full. Application of the definition of disconnectedness to the first pair gives:
δ(pianist, plays) = α(pianist) + α(plays) - 2 x α(FCH(pianist, plays)).
From Figure 2.8 can be seen that FCH(pianist, plays) is *plays*;
α(plays) = p and α(pianist) = p + q. Hence, δ(pianist, plays) = = (p + q) + p - 2(p) = q. Likewise, δ(pianist, beautifully) = α(pianist) + α(beautifully) - 2 α(FCH(pianist, beautifully)) = α(pianist) + α(beautifully) - 2 α(plays) = (p + q) + (p + s) - 2p = q + s. Since δ(pianist, beautifully) > δ(pianist, plays) it follows from the interpretation axiom that r(pianist, beautifully) < r(pianist, plays). The entire matrix of δ-values is given in Table 2.2. These δ-values imply the partial cohesiveness order over the

word pairs indicated by the graph in Figure 2.9, whose arrows are directed from higher to lower cohesiveness. It should be noted that DGL seems to work better where CGL shows its main shortcoming, viz. its failure to account for

Table 2.2 Matrix of δ-values, corresponding to the D-structure of Figure 2.8

	the	pianist	plays	beautifully
the	-	r	q + r	q + r + s
pianist		-	q	q + s
plays			-	s
beautifully				-

endocentricity in the data. As for the PWC (the, plays) vs. (pianist, plays), DGL yields a very plausible inequality prediction, in contrast with the

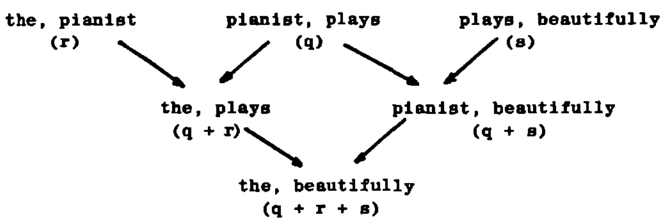


Figure 2.9 Graph representing the predicted cohesiveness order over the word pairs, given the D-structure of Figure 2.8

highly implausible equality prediction that would follow from application of CGL to the C-structure of Figure 2.9a.

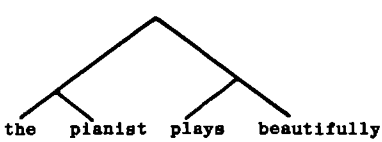


Figure 2.9a C-structure for the sentence *the pianist plays beautifully*

On the other hand, nothing in the formulation of DGL obliges us to set the value of the parameter r as less than that of q, a measure which would be necessary for preserving CGL's plausible prediction that r(the, pianist) exceeds r(pianist, plays). This would seem to place us in a dilemma; we shall devote further attention to this point in the last section of this chapter.

DGL's interpretation component can be formally characterized as relating cohesion judgments inversely to cost distances over the connectedness diagram implied by the D-structure. Some further explication will be useful. Firstly, the connectedness diagram, implied by a D-structure, is the graph resulting when its directed branches are changed into non-oriented *lines*. Each D-structure corresponds to exactly one connectedness graph. Conversely, however, the same connectedness graph may derive from as many D-structures as there are nodes in the diagram. These corresponding D-structures can be found by successively taking each of the nodes of the connectedness graph as the "root" and then changing all lines into branches descending from the root. Secondly, the notion of a cost distance (see Harary, Norman and Cartwright, 1965) implies that positive weights (costs) have in some way been assigned to the lines of the connectedness graph. The cost distance, then, between two nodes A and B is equal to the sum of the weights of the lines constituting that unique path connecting A and B.

We can illustrate this by reference to Figure 2.10, giving the connectedness diagram corresponding to the D-structure of Figure 2.8. Parameters representing the weights have been added to the lines of the graph.

$$\text{the} \;\overset{r}{\text{———}}\; \text{pianist} \;\overset{q}{\text{———}}\; \text{plays} \;\overset{s}{\text{———}}\; \text{beautifully}$$

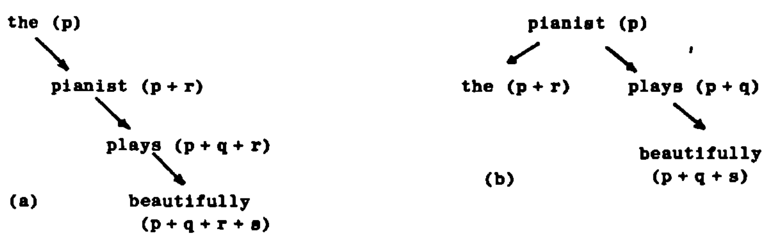
Figure 2.10 Connectedness graph for the sentence *the pianist plays beautifully* corresponding to the D-structure of Figure 2.8, with weighted lines.

The cost distance from *pianist* to *beautifully* is the sum of the weights along the path that connects the two words, hence equal to q + s. Likewise, d(the, plays) = r + q. The parameters have been chosen in such a way that the resulting cost distances match the degrees of disconnectedness in Table 2.2. In general, such matching will occur when, as in the above example, the weights w(i, j) assigned to the lines (i, j) in the connectedness graph, equal the absolute differences of the dependency values α(i) and α(j) in the D-structure. Conversely, given a connectedness graph with weights w(i, j) such as that in Figure 2.10, it is possible to reconstruct the set of contingent "dependency-valued" D-structures corresponding to it. To do so, we assign to the root of any corresponding D-structure an arbitrary initial dependency value, and then, starting from the root, we increase the dependency values along the paths in such a way that for all branches (i, j): $\alpha(j) = \alpha(i) + w(i, j)$.

The set of dependency valued D-structures, together with the D-structure of Figure 2.8 corresponding to the connectedness graph of Figure 2.10 -irrespective of their linguistic implausibility- is given in Figure 2.11. It should be clear that the D-structure in Figure 2.8 is the only linguistically plausible one.

This brief and informal consideration of the relationship between DGL and the notion of a cost distance, provides a convenient background for the discussion of two further issues raised by Levelt in connection with his dependency model.

Firstly, from the above considerations it becomes clear that in DGL it is the connectedness graph, implied in the D-structure, which determines the predicted cohesiveness order. These predictions are independent of the choice of the root, given a particular connectedness diagram. On the other hand, given an error free relatedness matrix, it is possible to derive that which the D-structures of Figure 2.8 and 2.11 share in common, viz. the connected-



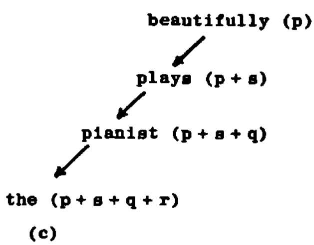


Figure 2.11 The three D-structures that, together with Figure 2.8, are compatible with the connectedness graph of Figure 2.10.

ness graph of Figure 2.10. But DGL does not enable us to use the r-values for determining the direction of the dependency relations or, for instance, the correct choice of root for the D-structure. "Only linguistic considerations", we read (Levelt, 1974c, p. 54), "can be decisive here, and not the relatedness data". A few lines further on Levelt continues: "Every linguistic dependency theory will indicate the element which is to be taken as the starting symbol. With the interpretation axiom, we suppose that that choice cannot be justified empirically, and this is decidedly a realistic point of view". Nevertheless, this point of view is not subscribed to in this study. Inspired by Gaifman's (1965) careful formal comparison of C-theory and D-theory to be dealt with in Chapter 3, in Chapter 4 we shall present an alternative interpretation theory for the D-grammar. According to this alternative interpretation the predicted r-values are not invariant under the choice of the root, given a particular connectedness graph. Hence, unlike Levelt, we suppose that an empirical justification of this choice on the basis of cohesion judgments is indeed possible.

The second issue concerns the fundamental question of whether we can formulate the necessary and sufficient condition to be placed upon a data matrix if it is to be representable by at least one connectedness diagram. This question has already been posed in connection with CGL, where it led to the formulation of the requirement of ultrametricity. But for DGL, such a critical property has not yet been found. Furthermore, it is doubtful whether such a concise necessary and sufficient condition for imposition upon the order of the word pairs in the case of a D-structure, can be found at all. We saw, however, that since it can be equated with a cost distance, the degree of disconnectedness is a distance metric (for this notion see Section 2.4.1). Accordingly, in any representation in terms of DGL, the triangular inequality $\delta(i, k) \leqslant \delta(i, j) + \delta(j, k)$ must hold for all words i, j, k in the sentence. This triangular inequality can thus be made a necessary, though not sufficient requirement for a data matrix to meet. This, however, would only make sense if we are dealing with measurements on a ratio level. Nevertheless, the attempt to check whether cohesion data satisfy the triangular inequality, though only a necessary minimum requirement for DGL, is definitely worthwhile. A negative result, for instance, would be very informative. It would not only disprove DGL, but also CGL (which, as we have seen, can easily be reformulated as a model including a distance function) together with all other attempts to grasp cohesion judgments in terms of distance models. In Section 2.4.1 we shall return to this important problem.

2.4 CRITICAL REVIEW OF LEVELT'S INTERPRETATION THEORIES

For our study of the structural adequacy of linguistic theories and grammars we have, in Chapter 1, adopted a methodological point of view that is inspired by Joan Bresnan (1978) and referred to as the "realistic approach". This point of view acknowledges the fact that linguistic structure has to be studied on the basis of its mainly indirect and "imperfect" manifestations in primary and secondary language behaviour. An instance of such an indirect and imperfect manifestation is the cohesion judgment. On the ground of this indirectness an instrumentalistic or operationalistic use of the cohesion judgment is abandoned. Cohesion judgments are looked upon as data to be described and explained. The adequacy of a linguistic theory, grammar or syntax should appear from its incorporability into one or more models of language behaviour. This adequacy may therefore reveal itself in the extent to which it functions as an indispensable component of a theory correctly accounting for the cohesion judgment.

In considering Levelt's interpretation theories from this methodological point of view, we can indicate several points on which they are open to criticism. These points constitute the focus of this concluding section. It will be useful to preface the argument with a recapitulation of the symbolism given in Section 1.2 according to which a theory of cohesion or relatedness judgments, R, is expressed as

(1) $R = F(G;\ e_1,\dots,e_k,\varepsilon)$.

As previously stated, in such an integrated theory we distinguish a grammatical component, G, and an interpretation component with its bridging function between G and the observations R. The interpretation component gives an account of both the "disturbing influences" (by explicitly incorporating them into the theory as additional determinants) and the measurement theoretical aspects of the theory as far as error and the nature of the function F is concerned. In terms of this symbolism, any proposal for a cohesion theory can be characterized as a combination of options with regard to G, the non-syntactic determinants e_1, ..., e_k, ε and F. In the case of CGL, for instance, the C-grammar has been chosen for G; the model is purely syntactic, i.e. $k = 0$, and as far as ε is concerned, the model is formulated deterministically in terms of the ideal errorless case. The nature of F has been treated in Section 2.3.1, where the structural properties of the C-grammar as reflected in R have already been specified. Referring to this interpretation as F_C, CGL can now be characterized as

(2) $R = F_C(CG)$.

CGL turned out to be inadequate as a theory for cohesion judgments. In particular, the restriction that elements of a constituent may not stand in relationships of different strengths to external elements, was systematically violated by cohesion judgments for PWCs involving endocentric groups. This was Levelt's reason for rejecting the C-grammar and investigating the D-grammar as a possible improvement over the C-grammar's shortcomings. In other words, he replaces (2) with (3) where F_D denotes Levelt's interpretation theory for the D-grammar:

(3) $R = F_D(DG)$.

At the end of Section 2.3.1 we mentioned that, logically, CGL's shortcomings must either be traceable to a possible inadequacy in its syntactic component or to inadequacies in the interpretation component (or both). Indeed, it is our view that several problems can be signalized in connection with this interpretation component that ought to be scrutinized before taking any decision to replace the C-grammar by the D-grammar.

In terms of the symbolism introduced above, these problems pertain to all of the three "options" determining a particular interpretation theory: the options with regard to F, ε and the e's. We shall handle these problems in the subsequent sections. Section 2.4.1 will be devoted to the nature of the mapping F. Arguments will be advanced to abandon distance models for cohesion. In Section 2.4.2 we shall advocate the employment of probabilistic rather than deterministic interpretations theories. Finally, in Section 2.4.3 we shall try to show that Levelt's argument for the rejection of the C-grammar is perhaps a result of the purely syntactic formulation of CGL. This formulation renders the C-syntax responsible for all of the variation in cohesion judgments, part of which, however, must be due to what should be regarded as non-syntactic factors. Our tentative suggestion is therefore for a revision of the interpretation theory as an alternative to the rejection of CGL.

2.4.1 Wholesale rejection of distance models for cohesion

In dealing with Levelt's rejection of CGL we mentioned the presence of reasons for dismissing its interpretation component rather than its linguistic component in the first instance. The main reason for this comes from the arguments for excluding distance models for cohesion judgments. If these arguments are valid, they obviously apply to DGL as well, since we have seen that both CGL and DGL are members of the family of distance models.

Moreover, we should have to conclude that cohesion and similarity, despite their formal comparability at the level of data theory, differ in nature to such a degree that distinct types of analyses would be required. Many kinds of similarities data have been successfully analyzed in terms of distance models, such as multidimensional scaling analyses or those types of cluster analysis which somehow incorporate the notion of a distance, e.g. Johnson's (1967) hierarchical cluster analysis. All these procedures would in principle have to be excluded for syntactic purposes if the notion of a distance should turn out to be inadequate for the representation of cohesion. In view of these far reaching consequences, the decision to exclude distance models should be approached with caution.

We shall therefore make a careful comparison of cohesion and similarity in terms of the postulates characterizing a distance function, assess their differences and determine whether these must really compel us to abandon distance models for cohesion.

Distance metric. Let X be a set of points. A distance metric over X is a function of the Cartesian product $X \times X$ into the real numbers which assigns to every pair, $x, y \in X$ a number $d(x, y)$, its *distance* satisfying the following postulates.

P1 : $d(x, y) = 0 \Leftrightarrow x = y$; that is, the distance from a point to itself is always zero and its distance to another point is never zero.

P2 : $d(x, y) = d(y, x)$; that is, distance is symmetric.

P3 : $d(x, y) \leqslant d(x, z) + d(y, z)$; that is, distances satisfy the triangular inequality.

The first postulate. In studies that apply distance models to *similarities data*, the empirical analogue of the first postulate is that an object cannot be more similar to another object than it is to itself. This interpretation is seldom put to the test; it should be regarded as one of the intuitive justifications for the decision to represent similarity by means of distance (see Bezembinder, 1971, p. 301). But this is less obvious in the application to cohesion judgments. The specification of syntactic relations is essentially a discription of how *different* words or phrases combine to form greater wholes. Hence, an empirical interpretation of P1 in the sense that "no word coheres more with another word than it does with itself" would be vacuous on intuitive linguistic grounds. We are now left with the question of whether this non-interpretability of P1 in terms of cohesion should discourage us from applying distance models to cohesion data. In this connection we ought to bear in mind that axiomatically formulated theories generally contain certain "technical assumptions". In the model these function as mathematically necessary building blocks which are needed for enabling the deductions we wish to make, but lacking any bearing on the basic idea of the theory. As a matter of fact, P1 is in no sense central to Levelt's interpretation theories. Accordingly, we might agree to regard P1 as a technical assumption, thereby deferring the question of the tenability of distance models for cohesion to other arguments.

The symmetry postulate. The second postulate requires distance to be independent of the order of the points. There are instances, however, where cohesion judgments exhibit asymmetries. An investigator seeking justification for the abandonment of distance models for cohesion might be tempted to place the burden of his argument on this discrepancy. But things are not that simple. In connection with this second postulate, again either of two points of view might be adopted. On the one hand, it might be decided to interpret this postulate as expressing an empirical claim. The obvious empirical interpretation would then be that every word x coheres as much with y as y coheres with x. Accordingly, one would collect and analyze data with an eye to possible violations of $r(x, y) = r(y, x)$. Were violations to be discovered, the notion of a distance would either have to be removed from the interpretation theory or seek accomodation in an adjusted version. On the other hand, there is the alternative approach by which no empirical interpretation is imposed on the second postulate. As with the first postulate it will then function merely as a mathematical cog in a model which is to be tested only with respect to other empirical claims. Investigators adopting this point of view typically collect or adjust their observations so as to satisfy P2 prior to the analysis. In this connection, Beals, Krantz & Tversky (1968, p. 130), discussing P2 in respect of asymmetric similarity judgments at the observational level, remark:

> "The similarity between x and y is the same as between y and x. This property specifies something about the way in wich the dissimilarity ordering should be establised. If asymmetries arise, they must be removed by averaging or by an appropriate theoretical analysis that extracts a symmetric dissimilarity index. (...) In other words, observed asymmetries must be handled by other techniques prior to the application of the geometric model" (in our case: graph theoretic distance model, E.S.).

From the first point of view one examines the metricity of cohesion data on the question of symmetry. From the second point of view one examines the metricity of cohesion data while taking its symmetry for granted. The latter doesn't of course resolve the symmetry issue, but bypasses it. Nevertheless it should be adopted only after careful scrutiny. The approach is sound provided that the asymmetries on the observational level are not so prominent as to render the precautionary measures for removing them an unacceptable means for juggling problems out of sight. So far, experience with cohesion data, especially with seven point scale procedures, does not seem to invalidate an approach from the second point of view. In a large majority of cases cohesion values that are given for word pairs are invariant under permutation of the words in the presentation. Nevertheless, the minority of cases that yield asymmetric cohesion judgments are interesting enough to encourage research along the lines of the first point of view. In the later chapters of this book attention will be given to asymmetric cohesion.

As far as we can see, Levelt's models reflect the second point of view. Nothing in his text suggests that symmetry of distance is deliberately hypothesized. Moreover, the empirical relatedness values used by the author as examples are typically the result of prior techniques for removing possible asymmetries by summation over subjects and orders of presentation. For this reason, as in the case of P1, we shall not make the symmetry postulate the critical issue in reaching a decision on the adequacy or inadequacy of the distance concept for cohesion.

The triangle inequality. Our doubts about the adequacy of distance models for cohesion mainly concern the third postulate, the triangle inequality. We shall start by giving an intuitive idea of our objections, prior to an attempt toward a theoretical elaboration, in order to facilitate an empirically supported critique. The triangle inequality states that, given three points, no interpoint distance exceeds the sum of the other two interpoint distances. The intuitive objections are easy to express in terms of an implication of this inequality: viz. that no interpoint distance can be smaller than the difference of the remaining two interpoint distances.

Experience with cohesion judgments shows that maximal cohesion or, equivalently, minimal incoherence seems to occur between those words exclusively constituting a phrase, as between *the* and *girls* in *the girls left*. Correspondingly, this strong cohesion in the data should be represented by a very small distance in the model. According to the triangle inequality, this small distance between *the* and *girls* simultaneously imposes an upper limit to the difference of the distances from *the* to *left* and *girls* to *left*. This difference, small as it necessarily is in the model, should correspond to a small difference in cohesion between *the* and *left* on the one hand and *girls* and *left* on the other. But this, as we saw in Section 2.3, is inconsistent with the observations usually given for pairwise comparisons involving endocentric constructions. We are therefore doubtful about the possibility of simultaneously expressing the very strong cohesion between *the* and *girls* and the great difference between the cohesion of (the, left) and (girls, left) in a distance model.

However, in the attempt to test the triangle inequality for cohesion judgments, we again find ourselves in a methodological dilemma. On the one hand Beals, Krantz & Tversky (op.cit.) point out that this property involves numerical addition, as a consequence of which it is not obvious how it is to be tested using ordinal data. On the other hand, if we would have unrelatedness measures $u(i, j)$ for all word pairs, i, j on a *ratio scale*, then we could write in the triangle inequality in terms of these u-values in the following way: $u(i, j) \leqslant u(i, k) + u(j, k)$; P3 would then in principle be testable. But, as Bezembinder (1970) points out, the question of whether we ought to impose this requirement on the data depends on whether the interpretation of the model endows it with an empirical claim. Now, Levelt's interpretation theories, though incorporating distance models, deduce ordinal predictions with respect to cohesion. Consequently, we have to decide on the status of the triangle inequality in these models. Is it to form part of their empirical interpretations, or, like P1 and P2, merely a formal property of an auxiliary mathematical concept necessary for enabling the interpretation theories to predict a particular cohesiveness order? In our view, in the case of P3 this latter viewpoint is untenable.

Firstly, if only the cohesiveness orders as they result from CGL and DGL were to be derived, then more parsimonious versions of the interpretation theories, without distance metrics, would suffice. CGL's predictions, for instance, could be stated directly in terms of the inclusion relation over the constituents. DGL's predictions can also be derived from an inclusion relation over paths: whenever the path from i to j is a subpath of that from k to l, then $r(i, j)$ exceeds $r(k, l)$. Secondly, Levelt even uses the very special and severe case of P3, the ultrametric inequality $u(i, j) \leqslant \max\,[u(i, k); u(j, k)]$ as a critical condition for the adequacy of CGL. Hence, the requirement that cohesion data should satisfy the triangle inequality even seems to take second place compared to Levelt's empirical claims. It should therefore be concluded that the triangle inequality is part of the empirical interpretation.

Testing the triangle inequality. In proposing a test for this empirical claim, we shall have to anticipate a finding of central importance to this study, one that will be discussed in full in Chapters 5 and 6. This finding is that the *word pairs* of a sentence are "Luce-scaleable" with respect to their relatedness. This means that it is possible -at least no counter evidence has been discovered from the sentences analyzed up to the time of writing- to assign numbers r(i, j) (relatedness values) to the word pairs (i, j) of the sentence which are, in a statistical sense, proportional to the empirical choice frequencies. An equivalent though conceptually more relevant formulation would be in terms of unrelatedness values u(i, j), which ought then to be inversely proportional to the choice frequencies.

The relevance to our problem is obvious. (a) We will use Luce's choice theory to assign to the *word pairs* unrelatedness values u(i, j) which, in being unique up to a multiplicative transformation, are values on a ratio scale. (b) If these values are obtainable in a statistically satisfying way, i.e. if they are acceptable in terms of goodness of fit, then as a following step it will be examined whether they qualify as distances on the question of the triangle inequality.

More specifically, the procedure is as follows. Imagine, an experiment performed in which N subjects are presented with all PWCs existing for an n word sentence. For every pair of word pairs (i, j) versus (k, l) we obtain the frequencies n(ij.kl) and n(kl.ij) = N - n(ij.kl) of those subjects who choose (i, j) $\gtrsim$ (k, l) and (k, l) $\not\gtrsim$ (i, j) respectively. In the first step we ask ourselves whether the choice frequencies can be accounted for by a model of the following form:

$$(2.1) \quad p(ij.kl) : p(kl.ij) = u(k, l) : u(i, j)$$

or, equivalently,

$$(2.2) \quad p(ij.kl) = \frac{u(k, l)}{u(k, l) + u(i, j)} \qquad \text{(for all words i, j, k, l in the sentence)}$$

In the case of a positive answer we would have an acceptable description, acceptable in the sense of goodness of fit, of the data in terms of unrelatedness values u(i,j) for *word pairs* on a ratio scale. As a second step we could then check all triples of *words* i, j, k to see whether the relevant u-values satisfy or violate the triangle inequality.

In order to answer the above question, two problems must be solved: the problem of estimating the u-values and the problem of testing the goodness of fit. We shall confine ourselves here to an outline of the steps of the estimation and testing procedures, leaving a detailed description of the data analysis to Chapter 5. Assuming that the N subjects are all taken from a homogeneous population operating under the same set of probabilistic constraints, we can express the probabilities of the empirically obtained frequencies n(ij.kl) for all PWCs in terms of the u-parameters by means of the binominal formula. The likelihood of the entire experimental result can be written as the product of all these binominal probabilities, and thereby also as a function of the u-values. The maximum likelihood method may then be applied to obtain the u-values for which that likelihood is maximum. On the basis of these u-values, the theoretical choice probabilities and consequently the expected frequencies can be derived and compared to the observed choice frequencies by means of a chi quare test.

We shall now give the results of the application of this procedure to the data of an experiment in which 49 subjects judged all 45 PWCs taken from the sentence *de man koopt een tas* (the man buys a bag). Since this experiment has been designed to test the probabilistic models to be developed in Chapter 5, the reader is referred there for details. The unrelatedness values are measures on a ratio scale and therefore uniquely determined except for an arbitrary unit of measurement. Accordingly, a word pair of moderate cohesion, viz.(man, tas), was assigned the u-value 1.0 to establish a unity: the others follow as the result of the estimation procedure. The u-values are given in Table 2.3.

Table 2.3 Luce generated unrelatedness values u(i, j) for the sentence *de man koopt een tas* (above diagonal); below the diagonal: values u'(i, j) = ln k . u(i, j) with k = 14.164 (see text).

	de	man	koopt	een	tas
de	-	0.07	2.49	4.73	4.90
man	0.05	-	.22	4.26	1.00*
koopt	3.56	1.12	-	1.04	0.27
een	4.21	4.10	2.69	-	0.08
tas	4.24	2.65	1.35	0.10	-

* value fixed at 1 as the arbitrary unit of measurement

From the u-values of Table 2.3 the corresponding choice probabilities p(ij.kl) and p(kl.ij) for every PWC (i, j) vs. (k, l) can be derived by means of Formula (2.2). Multiplication of its outcome by N, the number of Ss (in this case 49) yields the expected choice frequencies $\hat{n}$(ij.kl) and

$\hat{n}$(kl.ij) that are to be compared with the observed frequencies for calculating the PWC's contribution to the overall χ^2. For a numerical example we shall calculate the contribution to the overall χ^2 from the PWC (koopt, tas) versus (een, tas), which we shall denote as (K, T) vs. (E, T). Application of the formula (2.2) yields:

$$p(KT.ET) = \frac{u(E, T)}{u(E, T) + u(K, T)} = \frac{0.08}{0.08 + 0.27} = .22$$

and

$$p(ET.KT) = 1 - .22 = .78.$$

Multiplying the outcomes by N (= 49) yields:

$$\hat{n}(KT.ET) = 10.9 \text{ and } \hat{n}(ET.KT) = 38.1$$

The observed frequencies for this PWC were: 7 and 42 respectively. Hence, the contribution to χ^2 from this PWC is:

$$(7-10.9)^2/10.9 + (42-38.1)^2/38.1 = 1.80.$$

Summing the contributions of all 45 PWCs produces an overall $\chi^2 = 46.47$, ($df = 45 - 9 = 36$; $p > .10$). In these data there is no evidence against the scaling assumption (2.1).

Given this positive result, we now can check for all triples of words i, j, k in this sentence to see whether they satisfy or violate the triangle inequality. Let us consider as an example, the triple *man, koopt, tas*. From Table 2.3 it can be seen that u(man, tas), which equals 1, exceeds the sum of u(man, koopt) and u(koopt, tas) = .22 + .27 = .49 by a factor of 2. Likewise, it can be found that none of the $\binom{5}{3} = 10$ triples satisfies the triangle inequality. In some of the triples a small difference between the maximum u-value and the sum of the other two might suggest the desirability of a significance test, e.g. in the triples *de, een, tas* and *de, man, een*. In most triples, however, these differences are striking, e.g. *de, man, koopt, de, man, tas* and *man, een, tas*. We are therefore forced to conclude that unrelatedness values, though qualifying as an acceptable description of the data on a ratio level, definitely do not qualify as distances.

In connection with this analysis attention should be paid to a possible objection. It could be pointed out that the above procedure is in essence equivalent to:

1. computing an optimal representation of the data according to a model in which choice probabilities are *directly* proportional to relatedness (r):

$$(2.3)\quad p(ij.kl) = \frac{r(i, j)}{r(i, j) + r(k, l)}$$

2. discovering whether the reciprocals of these r-values, $r(i, j)^{-1}$, act like distance measures; in this case by checking for the triangle inequality. The objection might now be that Luce (1961, p. 155), in applying his choice theory to *similarity* judgments, already considers 1/v, i.e. his notational counterpart of our 1/r, untenable as a distance measure on other psychological grounds. He argues:

"If 1/v is a distance measure, in the usual sense, then $1/v(x, x) = 0$, so

$$p(x, y; x) = \frac{1}{1 + v(y, x) \;/\; v(x, x)} = 1^*$$

for any y however similar to x. Although it is probably unnecessary to cite data to convince the reader that this is wrong, they do exist..".

As an alternative to 1/v, Luce then proposes log 1/v as a distance measure, provided that the unit of v is chosen so that $v(x, x) = 1$ in order to ascertain that $d(x, x) = 0$ and $d(x, y) > 0$ for $x \neq y$.

Two points should be made in connection with this objection.

1. Luce's rejection of 1/v concerns its use in distance models for *similarity*. There, $d(x, x) = 0$ is naturally interpreted in terms of the maximal similarity considered to exist between an object and itself. In application of distance models to *cohesion*, $d(x, x) = 0$ could at best only form part of a technical assumption, since the idea of words cohering with themselves is vacuous. Accordingly, no empirical restrictions correspond to it, no cohesion judgments involving pairs (x, x) are ever asked for and no empirical objection could have a bearing on it.

2. Besides this, it is interesting to see that the alternatively proposed measure, log 1/v, in our notation log 1/r or log u, also fails to meet the triangle inequality, when applied to our cohesion data. To show this, we took the reciprocals $r = 1/u$ of the u-values to the right of the diagonal in Table 2.3. These r-values were transformed multiplicatively, $r' = k \times r$ to bring them into a range from 0 to 1. There are, however, no cohesion judgments involving word pairs (i, i) and consequently no r(i, i)-values that could serve to assess the multiplicative constant k according to the requirement that r'(i, i) should equal unity. We therefore used the "near-to-zero" unrelatedness u(de, man) and decided to choose k as to warrant a "near-to-unity" r'(de, man) viz. 0.95. Thereupon, since $\log 1/r' = -\log r'$, the negative

* p(x, y; x) stands for p(xx.xy) in our notation

natural logarithms were taken as the alternative "distance"-measures, yielding the u'-values that can be found below the diagonal in Table 2.3. From these values it can be seen that only two triples, *de, koopt, een* and *de, koopt, tas* meet the triangle inequality. Two violate it by a narrow margin: *de, een, tas* and *man, koopt, tas*; the others very distinctly. This is a notable finding since -log v (Luce's notational counterpart) does behave as a distance measure in many applications in psychology. For Luce (op,cit., p. 155) the reason for examining -log v as a possible distance measure was

> "... because there is evidence from other sources that the logarithm of the v-scale act much like the interval scales that arise in Fechnerian and Thurstonian scaling, and because these scales have, in one way or another, been treated as measures of distance...".

In summary we must conclude that *incoherence*, measured either as $u = 1/r$ or as $u' = \log 1/r$, fails to meet at least one of the critical properties of distance. As a consequence, we regard distance models as inadequate for the representation of cohesion.

2.4.2 Desirability of probabilistic interpretation theories

Levelt's interpretation theories are deterministic. They are formulated in terms of the ideal case and take no explicit account of random error or stochasticity. This is especially unsatisfactory in the case of CGL with its many equality predictions that can hardly be tested against real experimental data. Moreover, after many experiments the experience thus far acquired shows that the choice proportions obtained by counting over subjects, often deviate, *sometimes dramatically, from* 1 *or* 0. *For an example of this the* reader is reminded of the discussion of the triangle inequality. This experience is incompatible with the idea that the subjects in these experiments are all taken from a homogeneous population operating under the same set of deterministic constraints. But despite this lack of uniformity in cohesion judgments, as a rule a rapid convergence to an overall result is obtained from summation over subjects. Similarly, high correlations are generally found among individual subjects as well as between individuals and the overall result. For the moment these considerations suggest that the variation encountered in cohesion judgments is more likely to result from a homogeneous population operating under the same set of probabilistic constraints, than from a heterogeneous population of subjects operating under different sets of deterministic (or probabilistic) constraints. In view of this our preference will be for a probabilistic formulation of any interpretation theory of cohesion judgments.

In the pages already devoted to the question of whether cohesion judgments satisfy the triangle inequality we have suggested the lines along which such probabilistic formulations are to be here developed. The applicability of Luce's choice theory is an important though not yet linguistically relevant finding. It tells us that relatedness indices r can be assigned to the word pairs so that the choice probabilities for the PWCs can be described as directly proportional to the r-values.

A linguistically interesting question in connection with Levelt's interpretation theories would have been therefore whether these r-values can be assigned according to the restrictions of the interpretation theories, without too great a loss in the predictability of the choice probabilities from these "linguistic" r-values.

In the "unrestricted" application of Luce's model, the model equations are:

$$p(ij.kl) = \frac{r(i,\ j)}{r(i,\ j) + r(k,\ l)}$$

Replacing the r(i, j)'s with $\alpha(SCC(i,\ j))$ according to CGL these equations would read:

$$p(ij.kl) = \frac{\alpha(SCC(i,\ j))}{\alpha(SCC(i,\ j)) + \alpha(SCC(k,\ l))}$$

Replacing the r(i, j)'s with $\delta(i,\ j)^{-1}$, where $\delta(i,\ j) = \alpha(i) + \alpha(j) - 2\ \alpha(FCH(i,\ j))$, according to DGL, these equations become:

$$p(ij.kl) = \frac{\delta(i,\ j)^{-1}}{\delta(i,\ j)^{-1} + \delta(k,\ l)^{-1}} = \frac{\delta(k,\ l)}{\delta(k,\ l) + \delta(i,\ j)}$$

Again, the maximum likelihood method might be used for the parameter estimation and the chi square test in order to assess the goodness of fit.

2.4.3 Desirability of "integrated" interpretation theories

For the "realistic approach" that we have taken, the adequacy of a particular grammar G is a matter of its indispensability for one or more performance theories. This is not to insist that there need be one or more domains of empirical phenomena that can be exhaustively explained by incorporating G into an interpretation theory. Rather, it is required that there exist empirical domains that cannot be fully explained were we to omit G from a theory dealing with this domain. In other words, there should be phenomena whose explanation requires incorporation of G as a necessary though by no means sufficient condition. Our belief is that cohesion judgments constitute such a domain and, according to the realistic approach we have characterized a cohesion theory by means of the expression $R = F(G, e_1, \dots, e_k, \varepsilon)$. According to this expression among other things a cohesion theory must specify a demarcation between the syntactic and non-syntactic determinants of the judgment process. What is syntactic in an interpretation theory, and what is not, is determined by the syntactic formalism of the grammar to be interpreted, together with the interpreter's decision as to what formal properties are regarded as codetermining cohesion. Looking from this point of view, it is perhaps worth offering two admonitions to any investigator primary interested in the structural adequacy of grammars, and, thereby in the development of interpretation theories.

Firstly, any violations occurring during the testing of interpretation theories should be carefully screened to see whether they pertain to the syntactic or non-syntactic determinants of the judgment process. It is only on this basis that research can meaningfully proceed, focusing either on rejection or modification of G or on readjustment of the extraneous factors.

Secondly, in comparing different grammars one should realize that syntax may not be demarcated in the same way. As a consequence, what is syntactic in one grammar (i.e. what falls under G in the interpretation theory) may well be non-syntactic in another (i.e. comes under $e_1, \dots, e_k$). An empirical comparison of grammars should therefore be preceded by a careful theoretical comparison of their formal properties. Only thus we can discern in which respects the grammars are complementary and in which, if any, they comprise real alternatives so that the relevant empirical questions can be formulated.

Bearing these admonitions in mind, let us reconsider Levelt's rejection of the C-model as discussed in Section 2.3.1 and illustrated on Page 40 for the sentence *the girls left*. His rejection was based on CGL's failure to account for what is referred to as "endocentricity" in the data. The intuited difference in cohesion between *girls, left* and *the, left* violated the equality prediction resulting from application of CGL to the structure of Figure 2.7. In a pure C-model, however, notions such as "head" and "dependent" that might serve as a basis for the prediction of endocentricity are not defined. Hence, in an interpretation theory for the pure constituent model, the concept of endocentricity should not be regarded as a syntactic notion; it is not a variable in G, but in $e_1, \dots, e_k$. Consequently, CGL's failure to account for endocentricity should not be regarded as an argument for rejecting the C-grammar, but rather as an argument for readjusting the interpretation theory. This could be achieved, for instance, by introducing a factor like *semantic importance*. What we learn from this type of violation is that cohesion judgments cannot be accounted for exhaustively on the basis of the C-grammar. But this, as we have seen, is not a requirement in order for the C-grammar to be "realistic".

In view of this latter qualification we should not overlook the fact that CGL correctly predicts the inequalities (the, girls) $>_c$ (girls, left) and (the, girls) $>_c$ (the, left). Rejection of the C-grammar on account of the endocentricity argument may therefore also amount to a rejection of the basis for these valid predictions. Admittedly, the nature of the inadequacy of CGL provides good reasons for examining the D-grammar as an alternative. Application of DGL to the connectedness structure

the ———— girls ———— left

results in a correct prediction of the inequality (girls, left) $>_c$ (the, left). But, as we have mentioned on Page 43, nothing in DGL requires (the, girls) to be more cohesive than (girls, left). So although replacement of the C-model with the D-model provides us with a means for taking account of "endocentricity", at the same time we seem to lose the means for accounting for "constituency", at least as far as the interpretation of DGL is concerned. We append this latter reservation since, as will appear in Chapter 3, a D-grammar can also be interpreted to specify constituency, albeit to a limited degree. As a consequence, another interpretation of the D-grammar with a different demarcation of syntax might be considered and will, in effect, be given in Chapter 4.

For the moment, however, we may summarize this section by recapitulating the following points:

1. Neither the C-grammar nor the D-grammar seem to provide the basis for an exclusively syntactic interpretation theory, i.e. an interpretation theory without e's, yielding an exhaustive account of cohesion judgments. This, however, is not yet a sufficient argument for regarding these syntaxes as inadequate.

2. On the contrary, thus far both the C- and D-grammars qualify for the label "realistic" in as far as either model specifies certain formal notions on whose basis valid predictions with regard to part of the cohesion judgments seem to be derivable.

3. Depending on the different delineations of syntax in the C- and D-gram-

mars, these predictions will affect different sets of PWCs which only partially overlap. It is only in as far as these sets intersect that a direct comparison of "syntactic" predictions is possible.

In order to obtain an idea of whether the predictions that are syntactic according to both interpretation theories are able to contradict each other, a careful theoretical comparison of their formal properties will be made in advance. This comparison comprises the content of Chapter 3. On the other hand, the predictions for PWCs which are syntactic for only one of the interpretation theories are not without interest from the "realistic" viewpoint. Whenever the syntactic prediction from either of the interpretation theories holds, this imposes a requirement on the competing interpretation theory, that it provide a succesful non-syntactic account for the selfsame cohesion judgment.

CHAPTER 3 / FORMAL COMPARISON OF CONSTITUENT THEORY AND DEPENDENCY THEORY

3.1 INTRODUCTION

A suitable preparation for an empirical comparison between C-theory and D-theory should start from a formal juxtaposition. It must first be established in which respects the theories agree and in which they differ, before truly relevant empirical questions can be posed. We shall take advantage of the contributions of Hays (1964), Gaifman (1965) and Fitialov (1973), among others, to the discussion concerning the equivalence of these rivaling theories.

In the discussion two facets can be distinguished. Firstly, the comparison may focus attention on either *linguistic theory* or *grammar*. Secondly, the discussion may be about the issues of *weak* or *strong equivalence*. As far as the first facet is concerned, grammars are particular options drawn from general possibilities specified by formalisms of grammatical theories. Both the given options and the general possibilities may be the objects of comparison. As for the second facet, a grammar G weakly generates a language L and strongly generates a system of structural descriptions Σ. Likewise, a theory's weak generative capacity is the collection of languages generated by the grammars provided by the theory, and its strong generative capacity is the collection of Σ's generated by the grammars according to the theory. Both these weak and strong aspects may be the object of comparison.

The above mentioned authors are mainly concerned with the two questions of whether C- and D- *theory* are weakly equivalent, and whether they are strongly equivalent. The answer to the first question is affirmative. For the proof of the weak equivalence, a matter with which this thesis is only peripherally concerned, the reader is referred to Gaifman (op.cit., p. 335 ff.). His theorem implies that the collection of languages generated by the D-formalism is equal to the collection of languages generated by the context free C-formalism. If the requirement for strong equivalence were to be formulated analogously to the requirement for weak equivalence, so that it would be the equality of collections of Σ's at issue, then the answer to the second question could only be negative. C- and D-theory elucidate different aspects of linguistic structure in a way which differs formally.

Hays, Gaifman and Fitialov, however, take a different approach toward the problem of comparing structural features of the rivaling theories. In a sense to be explained later, they reduce both formalisms to the same denominator in order to achieve a comparison of their strong generative capacities. This reduction to the same denominator amounts roughly to the following procedure: (i) a priori definitions of certain types of "correspondence" between a C- and D-structure are given; (ii) the requirement for C- and D-theory to be strongly equivalent (Hays uses the term "strongly equipotent") is weakened: the structures assigned to strings by grammars of the two types need not be the "same" but must "correspond" in the sense of (i); (iii) the strong equivalence under each of the distinguished types of correspondence is asserted or denied.

Before giving a more detailed report of this global procedure, certain decisions must be made with respect to the notions, terminology and notation to be used. In these respects there are considerable differences among the authors referred to above. In what follows, therefore, we shall attempt a unified summary of the equivalence discussion by means of a *set-theoretical* representation of sentence structure to be introduced in the next section. This description in set-theoretical terms is in the first instance intended to simplify a rather complex discussion. Besides, it will provide a good starting point for our presentation of the interpretation theories in the next chapter, which are themselves of a set-theoretical nature.

3.2 SET-THEORETICAL REPRESENTATION OF SENTENCE STRUCTURE

In the following, the sentences generated by a grammar will be considered interchangeably as sequences of words or as sets of word occurrences. Strictly speaking, word occurrences (see Gaifman, 1965, p. 305) are ordered pairs (w,p), w being the word type and p the place number in which the word occurs. Hence, a word sequence, s, may be uniquely determined from the set of word occurrences and vice versa. No confusion will result from continuing to denote the word occurrences as $i,j,k...$ as was done in the previous chapter.

Let W be the set of word occurrences of a sentence s. We choose the symbol W^* to denote the collection of all subsets of W. The elements, C, of W^* will be called *clusters*. For instance, let W be $\{i,j,k\}$, then the set of clusters is $W^* = \left\{\emptyset,\{i\},\{j\},\{k\},\{i,j\},\{i,k\},\{j,k\},\{i,j,k\}\right\}$. Now, let us suppose that the sentence s has been generated by a particular C-grammar and that a particular C-structure S has been assigned to s. Part of this structural description is the specification of how the sentence s is partitioned into substrings of a specified constituent type, how these substrings are partitioned further, and so on, until the terminal elements, the word occurrences, are reached. Since each of these substrings is itself a subset of W, it is possible to represent the C-structure S set-theoretically as a particular subset of W*: $S \subset W^*$. For instance, let s be $\{i,j,k\}$ and the C-structure assigned as follows:

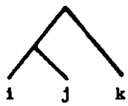


Figure 3.1

then $S = \Big\{\{i\},\{j\},\{k\},\{i,j\},\{i,j,k\}\Big\} \subset W^*$

If, however, the sentence s has been generated by a D-grammar, then the type of structural information assigned to s by this grammar seems to be of an entirely different nature. In the literature on the formal comparison of C- and D-grammars, however, some formal notions are defined for D-structures that are comparable to, though not the same as constituents. These notions will now be introduced with reference to an example. Given the D-tree in Figure 3.2

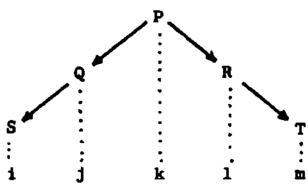


Figure 3.2

one can illustrate the following two notions: (a) complete subtrees, (b) subtrees.

(a) A *complete subtree* of a given node N is the node N together with all other nodes directly or indirectly dependent on N.

Examples

The complete subtree of node R is:

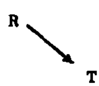


Figure 3.3

The complete subtree of node Q is:

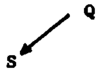


Figure 3.4

The complete subtree of node P is the D-tree of Figure 3.2 itself.
It can be shown that each node defines just one complete subtree.
(b) A *subtree* of some node N is the remainder of the complete subtree of N after deleting from it k directly dependent nodes together with their complete subtrees, where $0 \leq k < n$, n being the number of direct dependents of N. Each node has one or more subtrees.
Examples
One subtree of P is:

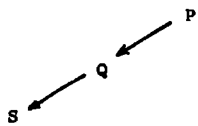


Figure 3.5

It is obtained from the complete subtree of P by pruning from it the complete subtree of R.
Another subtree of P is:

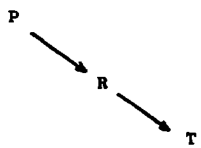


Figure 3.6

Here, the subtree of Q is pruned away.
Furthermore, all complete subtrees are by definition special cases of subtrees.

When lexical insertion takes place, single word occurrences are assigned to each of the nodes of the D-tree. Hence, on the basis of this lexical insertion, it is possible to associate the subtrees and complete subtrees with sets of word occurrences in a way analogous to that done for the C-structures. If we call the sets of word occurrences associated with the subtrees *substructures*, and, by definition, add single word occurrences to the set of substructures, it becomes possible to define set-theoretically for the D-structure in Figure 3.2:

(a) the set of *complete substructures*:

$$S_1 = \left\{\{i\},\{j\},\{k\},\{l\},\{m\},\{i,j\},\{l,m\},\{i,j,k,l,m,\}\right\},$$

(b) the set of *substructures*, S_2:

$$S_2 = S_1 \cup \left\{\{i,j,k,\},\{k,l,m\}\right\}$$

Note that both S_1 and $S_2 \subset W^*$

So both C-theory and D-theory can be said to specify a particular subset from the collection of all subsets of word occurrences.

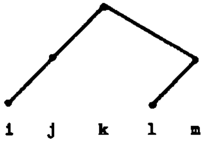
3.3 THE DISCUSSION OF STRONG EQUIVALENCE

As already mentioned, in the discussion of equivalence between C- and D-grammars attention becomes focused on the question of whether a one-one mapping can be found between C- and D-grammars such that they may be said to *correspond* in a certain previously defined manner. In this equivalence discussion three definitions of correspondence are involved. In the following, these will be referred to, quite neutrally, as type 1 correspondence and, in view of the relationship of the others, as type 2a and type 2b correspondence.

Type 1 correspondence

A C-structure and a D-structure are in type 1 correspondence if every *constituent* in the C-structure equals a *complete substructure* in the D- structure and if every *complete substructure* in the D-structure equals a *constituent* in the C-structure.

This type of correspondence holds in the following example:



versus

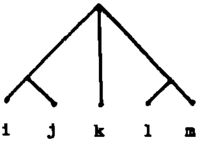
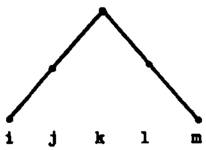


Figure 3.7a

Figure 3.7b

because both structures specify the same subset of W^*. For reasons of convenience, this set is given in the simplified form of a collection of substrings: {i, j, k, l, m, ij, lm, ijklm}, a simplification which will henceforth be employed more often. Type 1 correspondence also holds between the members of the following pair of structures:



versus

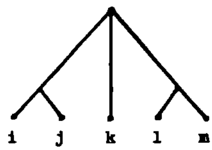
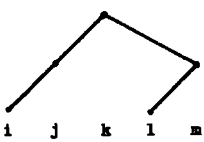


Figure 3.8a

Figure 3.8b

for the same reason. But it does not hold in the following case:



versus

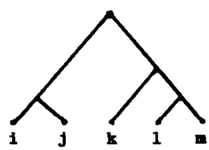


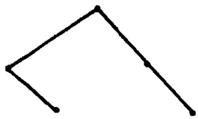
Figure 3.9a

Figure 3.9b

the reason being that the subset of W*, specified by the structure of Figure 3.9b, contains {k,l,m} as a constituent which is not a complete substructure in the D-structure of Figure 3.9a.

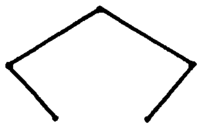
This type of correspondence is not mentioned in Hays' (1964) paper. In Gaifman's article (p. 316) the set of complete substructures of a D-structure is called "the phrase structure ramification induced by the D-structure". In the English translation of Fitialov's (1973, p. 134) contribution this type of correspondence is called *full correspondence*.

Figures 3.7a to 3.9b inclusive suffice to illustrate the main findings of the discussion of the strong equivalence of C-theory and D-theory under type 1 correspondence. These points are as follows: (i) under type 1 correspondence for every D-structure it is possible to construct exactly one corresponding C-structure. The proof, too lengthy to be repeated here, is due to Gaifman (op.cit., p. 334 ff.). (ii) Different D-structures may give rise to the same C-structure. This is the case for the D-structures given in Figures 3.7a and 3.8a which both are in type 1 correspondence with the C-structure in Figure 3.7b (or 3.8b). Likewise, the D-structures



i j k l m Figure 3.10

and



i j k l m Figure 3.11

are also in type 1 correspondence with the C-structure in Figure 3.7b.

The relation of type 1 correspondence is therefore unambiguous only in one direction. Only one C-structure goes with a given D-structure, but several D-structures may go with a given C-structure. (iii) There exist C-structures for which no type 1 corresponding D-structure can be constructed. This is clearly illustrated by the pair of structures given in Figures 3.9a and 3.9b. The class of C-structures that can be put in type 1 correspondence with D-structures constitutes a trivial subclass of the class of C-structures. Fitialov (op.cit., p. 128) couches it in plain terms:

> "The IC-structures (C-structures in our terminology) thus obtained from the dependency structures have one essential feature in common, namely, every constituent of more than one element has a single element IC, this is the main element of the group ("complete substructure" in our terminology) in the dependency structure. Such IC-structures will henceforth be called simple structures."

This property, shared by all those C-structures potentially in type 1 correspondence with a D-structure, has been formulated more exactly by Gaifman (op.cit., p. 322). In his terms only C-grammars of degree 0 or 1 can be placed in type 1 correspondence with D-grammars. The degree of a grammar is defined as follows:

Definition (degree): Let G be a C-grammar. A category X of G is of degree 0, to be denoted "deg(X)=0" if X does not appear on the left side of any rule $\alpha \rightarrow \beta$ (i.e. the only strings of words of category X are single words). Deg(X)=n if deg(X)$\neq$0,1,...,n-1 and for every rule of the form $X \rightarrow \beta$, ($\beta = Y_1,\ldots,Y_k$) there is a Y_i, $1 \leq i \leq k$, such that deg(Y_i)=n-1. X is of infinite degree, deg(X)=∞, if for no n deg(X) = n. The degree of the grammar deg(G)= max deg(X), with X ranging over all categories of G.

Example

If the rules of G are S $\rightarrow$ AB, A $\rightarrow$ ij, A $\rightarrow$ lm, B $\rightarrow$ kA, yielding the C-structure of Figure 3.9b (after appropriate labeling of the non-terminal nodes), then: deg(i) = deg(j) = deg(k) = deg(l) = deg(m) = 0, deg(A) = deg(B) = 1, deg(S) = 2 and so, deg(G) = 2.

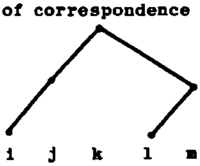
Because the degree of this grammar exceeds 1, it is impossible by means of a D-grammar to imitate what this C-grammar states about the C-structure of the string ijklm, where "imitate" means "imitate under type 1 correspondence".

Hence, if the question of strong equivalence has to be answered on the basis of the relation of type 1 correspondence, then the apparent conclusion must be in the negative. Strong equivalence, then, is only possible for C-grammars whose degrees do not exceed unity.

Type 2a correspondence

A C-structure and a D-structure are in type 2a correspondence if every *constituent* in the C-structure equals a *substructure* in the D-structure and if every *complete substructure* in the D-structure equals a *constituent* in the C-structure.

This type of correspondence holds in the following example:



versus

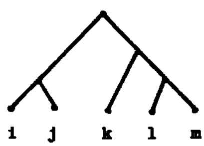


Figure 3.12a

Figure 3.12b

This is easily verified: the set of complete substructures in the D-structure of Figure 3.12a is:

$$S_1 = \{i,\ j,\ k,\ l,\ m,\ ij,\ lm,\ ijklm\}\ ;$$

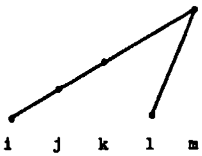
the set of substructures in the same D-structure is:

$$S_2 = S_1 \cup \{ijk,\ klm\}$$

and the set of constituents in the C-structure of Figure 3.12b is:

$$S_3 = S_1 \cup \{klm\}.$$

But it does not hold between the members of the following pair:



versus

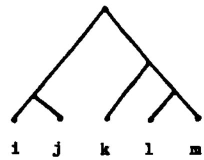
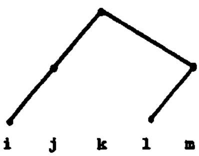


Figure 3.13a

Figure 3.13b

because the constituent klm in Figure 3.13b is not a substructure in Figure 3.13a. Again it holds in:



versus

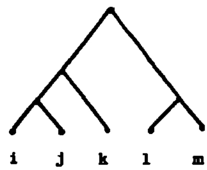
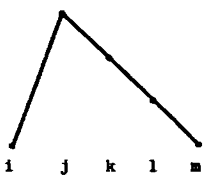


Figure 3.14a

Figure 3.14b

and in the following couple of examples:



versus

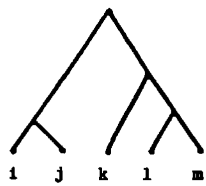


Figure 3.15a

Figure 3.15b

This type of correspondence, called relational correspondence by Hays (op.cit., p. 520) and strong correspondence in the English translation of Fitialov's (op.cit., p. 134) paper, is ambiguous in both directions. The examples 3.12a to 3.15b inclusive, suffice to show this. While the D-structures of Figures 3.12a and 3.14a are the same, their C-structural counterparts differ. On the other hand, it can be seen from Figures 3.12b and 3.15b that one and the same C-structure may correspond to different D-structures.

It is obvious that whatever the answer to the question of equivalence under type 2a correspondence might be, it can hardly be of any relevance to our central problem. For, if the answer were positive, it would be based on an idea of correspondence which is much too flexible. Under type 2a correspondence both the D-grammar yielding the D-structure of Figure 3.12a and the D-grammar yielding the D-structure of Figure 3.15a might be equivalent with a C-grammar yielding the C-structure of Figure 3.12b (or 3.15b). But is is clear that these two D-structures represent quite different assertions about the structure of the string ijklm. Besides the fact that in the two cases different elements assume the role of the center in the sentence, not even the connectedness networks are the same. This argument can be strengthened by the consideration that the C-structure of Figure 3.12b results from four bifurcations, for each of which D-theory would characterize one of the fork's prongs as the head of the construction. The resulting D-structure would be one of $2^4 = 16$ possible D-structures, all of them essentially different but nevertheless in type 2a correspondence with the C-structure of Figure 3.12b. Two of these sixteen examples can be used to complete the argument, viz.:

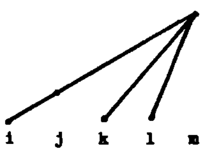


Figure 3.16

and

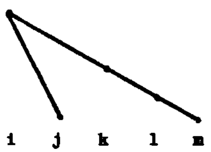


Figure 3.17

Hays, as a matter of fact, is the first to admit the latitude of his notion of relational correspondence. This latitude results from the fact that, until now, the labeling properties of both formalisms have been left out of the equivalence discussion. Hays (op.cit., p. 520 *ff.*) introduces with approval Gaifman's restriction on labeling -to be specified hereafter- in order to avoid "counterintuitive matches" between the formalisms. He illustrates this by means of the examples given in Figures 3.18a to 3.18d inclusive, where two C-structures and two D-structures are given for the ambiguous sentence *They are flying planes.*

are flying as verbal part of the verb phrase

flying planes as nominal part of the nominal predicate

C-structure :

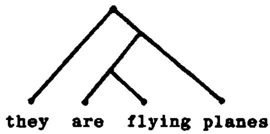


Figure 3.18a

Figure 3.18b

D-structure :

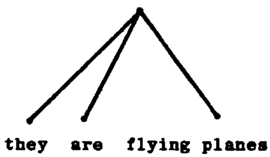


Figure 3.18c

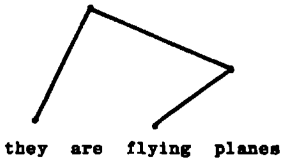


Figure 3.18d

The structures given in Figures 3.18a and 3.18c are in type 2a correspondence, because all complete substructures in the latter are constituents in the former and all constituents in the former are substructures in the latter. For the same reason the structures in 3.18b and 3.18d correspond "relationally", whereas those in Figures 3.18a and 3.18d do not: *are flying* is a constituent in Figure 3.18a without being a substructure in Figure 3.18d. But the counterintuitive match signalized by Hays is the type 2a correspondence holding between the structures of Figures 3.18b and 3.18c !

It is the function of the already announced type 2b correspondence, that will be defined presently, to cope with this latitude problem. In preparation for this definition it is useful to give first a somewhat free extension of the notion of weak generative capacity, so as to render it applicable not only to formalisms or grammars, but to the auxiliary symbols within a grammar as well. Both in C- and in D-theory strings derive from auxiliary symbols. In C-theory these symbols are the labels of the constituents covering the (sub)strings involved. In D-theory these symbols label the heads of the complete subtrees covering the (sub)strings. The total set of (sub)-strings derivable from an auxiliary symbol will be called the weak generative capacity of that symbol.

Type 2b correspondence

A C-structure and a D-structure are in type 2b correspondence if: (i) they are in type 2a correspondence and (ii) every *complete subtree* in the D-structure has a head, labeled with an auxiliary symbol with the same generative capacity as the auxiliary symbol labeling the co-extensive *constituent* in the C-structure (i.e. Gaifman's restriction on labeling).

This type of correspondence no longer holds between the structures in Figures 3.18b and 3.18c. On the one hand, the D-grammar generating the D-tree of Figure 3.18c will have to label the "complete subtree" accounting for the substring *are* with an auxiliary symbol having the weak generative capacity of a certain subclass of auxiliary verbs. On the other hand, the C-grammar generating the C-tree of Figure 3.18b will have to use a symbol to dominate *are* with the weak generative capacity of a copula. Hence, the auxiliary symbols "accounting for" *are* possess differing weak generative capacities.

Gaifman (op.cit.) shows the strong equivalence of the formalisms under type 2b correspondence when infinite C-grammars are excluded. The great latitude characteristic of type 2a correspondence has been considerably reduced by the labeling restriction. However, the type 2b correspondence thus obtained is still not restrictive enough. Four minigrammars, two of the dependency type, DG_1 and DG_2, and two of the constituent type, CG_1 and CG_2, suffice to illustrate this.

DG_1 consists of the following rules:

(i) *(V) (ii) V(N,*,N) (iii) N(A,*) (iv) A(*)

and the lexical assigning rules:

V/hears, V/sees, N/man, N/woman, A/a, A/the.

DG_2 consists of the following rules:

(i) *(V) (ii) V(A,*,A) (iii) A(*,N) (iv) N(*)

and the same set of assigning rules as in DG_1.

CG_1 consists of the rules:

(i) S → NP+VP (ii) VP → V+NP (iii) NP → A+N
(iv) V → {hears,sees} (v) N → {man,woman} (vi) A → {a,the}.

CG_2 equals CG_1 except for the rules (i) and (ii). These are replaced by:

(i) S → KP(kernel phrase) + NP (ii) KP → NP+V.

All these grammars are weakly equivalent, their weak generative capacity being the set of 32 sentences resulting from five successive choices from the following sets: {the,a}, {man,woman}, {sees,hears}, {the,a}, {man,woman}. The structural differences between the grammars become clear upon derivation of the structures which they respectively assign to the sentence *the man hears the woman*.

DS_1 generates the structure:

DS_1 =

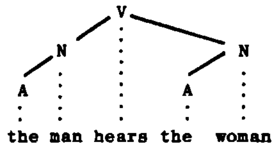


Figure 3.19

DS_2 generates the structure:

DS_2 =

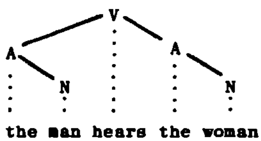


Figure 3.20

CG_1 generates the structure:

CS_1 =

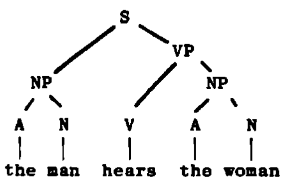


Figure 3.21

CG_2 generates the structure:

CS_2 =

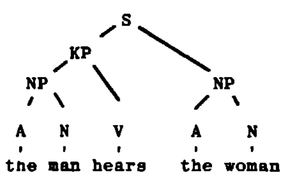


Figure 3.22

It can be easily verified that type 2b correspondence holds between (a) all DG_1 structures and all CG_1 structures; (b) all DG_1 structures and all CG_2

structures: (c) all DG_2 structures and all CG_1 structures and (d) all DG_2 structures and all CG_2 structures. The first of these assertions will now be demonstrated, the others may be checked in the same way.

First it will be shown that DS_1 and CS_1 are in type 2b correspondence. As far as the first requirement for this type of correspondence is concerned, namely that the structures should be in type 2a correspondence, the reader is referred to the type 2a correspondence existing between the structures of Figures 3.12a and 3.12b, which are co-extensive with DS_1 and CS_1 respectively.

In order for the second requirement (label correspondence) to hold it is necessary and sufficient to ascertain the points (a) to (c).

(a) The complete subtrees in DS_1 are those covering the (sub)strings 1. *the man hears the woman*, 2. *the man*, 3. *the woman*, 4. *the*$_1$ (i.e. the first occurrence of *the* and 5. *the*$_2$.

(b) In DS_1 the heads of these subtrees are respectively: 1. V. 2. N, 3. N, 4. A, 5. A. In CS_1 the auxiliary symbols dominating the coextensive constituents are: 1. S, 2. NP, 3. NP, 4. A, 5. A.

(c) As for the weak generative capacities (WGCs) of these symbols, the following equalities hold: WGC(V) in DG_1 = WGC(S) in CG_1 (namely both equal the WGC of the grammars themselves). WGC(N) in DG_1 = WGC(NP) in CG_1 (namely the set of 2 x 2 = 4 substrings resulting from the successive lexical choices from the sets {the,a} and {man, woman})$\cdot$WGC(A) in DG_1 = WGC(A) in CG_1 = {the,a}.

This demonstration of type 2b correspondence is invariant under lexical insertion. Therefore, if for the strong equivalence of both grammars it is required that they assign type 2b corresponding structures to the same sentences, this equivalence is established.

It should be noted that the VP in CS_1 does not play any role in the assessment of the type 2b correspondence. Neither would KP in CS_2 exclude type 2b correspondence between this structure and DS_1 . Hence, with regard to the strong equivalence under type 2b correspondence, the same remarks as those concluding the section on type 2a correspondence apply, albeit to a more limited degree. Again the equivalence is based on a notion of correspondence which permits too much latitude. Under type 2b correspondence, both CG_1 and CG_2 are strongly equivalent with DG_1, whereas it is clear that these C-grammars make quite different assertions about the structure of sentences such as *the man hears a woman*. There is no information available in the corresponding D-structure enabling us to choose between one of the C-grammars. By their very nature, D-structures fail to indicate differences between degrees of relationship existing between the V and the first N and between the V and the second N. C-theory simply possesses extra properties which cannot be imitated by a D-grammar. The same remarks apply in the opposite direction. D-theory has extra properties which a C-theory lacks. In the C-structures CS_1 and CS_2 nothing enables us to select either DS_1 or DS_2 as the most naturally corresponding D-structure, because there is no indication of whether N governs A or the other way round. What we see here is that the strong equivalence under type 2b correspondence becomes established on a level of abstraction where even quite different things may seem alike. For this reason, the strong equivalence proof does not detract from the sense of attempting to compare the structural adequacies of the rivaling theories.

Some additional remarks should be made on the equivalence discussion presented thus far.

(i) From a formal point of view there is of course no limit to the number of conceivable types of correspondence under which the equivalence question might be posed. Moreover, other types of correspondence might perhaps throw a different light on the comparison of the formalisms involved.

(ii) The nature of the equivalence discussion reflects the fact that, historically, C-grammars preceeded D-grammars. Accordingly, the questions asked tend to be more of the type "Are D-grammars able to imitate what C-grammars do?" rather than the other way round. Less attention is paid to the issue of dependency than to that of constituency. And, as far as constituency is concerned, there is a bias towards viewpoints favouring C-grammars. Let us consider the latter statement in some detail with reference to the D-structure DS_1 in Figure 3.19 and the C-structure CS_1 of Figure 3.21, which are in type 2b correspondence. Seen as a set of complete substructures DS_1 is a subset of CS_1; $CS_1 = DS_1 \cup \{\{\text{hears, the, woman}\}\}$

But, if we regard DS_1 as a set of substructures, then CS_1 is a subset of DS_1 with its overlapping substructures; $DS_1 = CS_1 \cup \{\{\text{the, man, hears}\}\}$

Hence, in requiring of the D-grammar that it set a *complete substructure* against every *constituent* in the C-structure, and from the C-grammar that it set a *constituent* against every *complete substructure* in the D-structure (type 1 correspondence), is to render the class of D-grammars a trivial subclass of the class of C-grammars. In requiring of the D-grammar that it set a *substructure* against each *constituent* of a C-structure, and from the

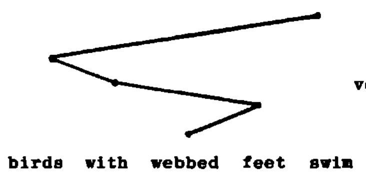
C-grammar that it set a *constituent* against each *complete substructure* (type 2a and type 2b correspondence) represents less of a demand on D-grammars, but induces the ambiguities discussed above.

But a third possibility, requiring a one-to-one correspondence between *substructures* and *constituents*, never comes up for consideration in the equivalence discussion. Obviously, none of the authors have been willing to interpret the possibility of specifying substructures and overlapping substructures as an extra property of D-grammars which is absent from C-grammars. But there is no compelling reason for neglecting this last mentioned property of D-grammars. On the contrary, this property might appear (and in fact, will appear) to be of empirical importance. These considerations bring us to the introduction of a third type of correspondence.

Type 3 correspondence

A C-structure and a D-structure are in *type 3 correspondence* if every *constituent* in the C-structure equals a *substructure* in the D-structure and if every *substructure* in the D-structure equals a *constituent* in the C-structure.

This type of correspondence holds between the following structures:



versus

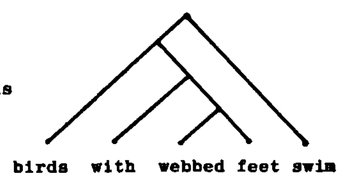
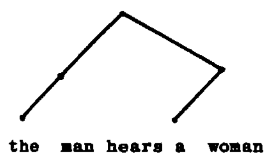


Figure 3.23a

Figure 3.23b

But is does not hold between the structures:



versus

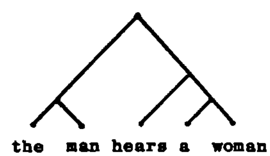


Figure 3.24a

Figure 3.24b

All constituents in structure 3.24b are substructures in structure 3.24a, but the substructure covering *the man hears* in the latter is not a constituent in the former.

As soon as a D-grammar exhibits a rule assigning more than one dependent to a governor, we are confronted with overlapping substructures and with the impossibility of type 3 correspondence to any C-grammar at all. So if the question of strong equivalence among the competing formalisms were to be answered on the basis of type 3 correspondence, the answer would be negative. Strong equivalence, in that case, would only hold between a trivial subclass of D-grammars and a trivial subclass of C-grammars. These D-grammars consist exclusively of rules assigning at most one dependent to a governor: X(Y,*), X(*,Y) or X(*). The C-grammars in the trivial subclass only embrace rules of the form $A \rightarrow a(\varphi)$ or $A \rightarrow (\varphi)a$ where *a* is an obligatory terminal element and φ at most one element of the non-terminal or terminal vocabulary. Consequently, under type 3 correspondence, the D- and the C-formalisms essentially differ.

3.4 CONCLUSIONS

Let us agree to denote the (sub)strings of word occurrences of a given sentence covered by such constructs as constituents, complete substructures and substructures, with the general term: *syntagmas*. Then both C-grammars and D-grammars may be viewed as making formal assertions about how the word occurrences of a given sentence combine to form syntagmas. This specification of a set of more and more inclusive syntagmas can be conceived of as the specification of a subset of W^*, in the set-theoretical terminology introduced in Section 3.2. From the equivalence discussion, it must be concluded that the competing formalisms differ essentially in the ways they specify these syntagmatic structures. Syntagmatic structures corresponding to C-grammars whose degree exceeds unity cannot be imitated by D-grammars; the non-hierarchical syntagmatic structures corresponding to the D-grammar's substructures cannot be imitated by C-grammars.

So far, the syntagmas are only formal notions. It is of course an empirical question whether they play a role in the process of making cohesion judgments. And if so, it is again en empirical question whether these syntagmas are identifiable as the set of constituents assigned by a C-grammar, or as the set of substructures, complete or otherwise, assigned by a D-grammar.

Let us be more concrete about this by making use of the example that will play an important role throughout this thesis. For this purpose we choose a sentence of the type: art_1 $noun_1$ $verb_{transitive}$ art_2 $noun_2$, small enough to be experimentally manageable, but sufficiently large to allow for a demonstration of the central issue. The current structure assigned to sentences of this type (e.g. *the man hears the woman*) by C-theory is:

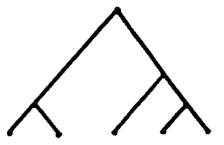


Figure 3.25

being a structure of degree > 1.

The current structure assigned by D-theory is:

Figure 3.26

where a D-rule, assigning more than 1 dependent to a head-category is involved: V(N,*,N)

In other words, these structures exhibit those critical features which render impossible a strong equivalence under type 1 or type 3 correspondence (the only ones without unacceptable latitude).

Put in set-theoretical terms, the empirical question is now as follows. Is it possible to make an empirically based choice between the following sets of syntagmas (where, for convenience, the words are reduced to their initials):

CS_1 = {T,M,H,A,W,TM,AW,HAW,TMHAW}

and (using only complete substructures):

DS_1 = {T,M,H,A,W,TM,AW,TMHAW}

and (using all substructures):

DS_2 = {T,M,H,A,W,TM,AW,TMH,HAW,TMHAW} ?

Furthermore, if we turn from the level of comparing grammars and consider rather the level of comparing formalisms, then our attention must become broadened to include certain other options as well. Among the options proffered by the C-formalism the following might be of interest:

$$CS_2 = \{T,M,H,A,W,TM,AW,TMH,TMHAW\}$$

in which the verb combines with the first noun phrase, or:

$$CS_3 = \{T,M,H,A,W,TM,AW,TMHAW\}$$

in which the C-grammar is of degree 1 and hence in type 1 correspondence with the D-grammar yielding DS_1 (hence, $CS_3 = DS_1$).

Among the options proffered by the D-formalism, structures differing from DS_1 and DS_2 by a different choice of the central element, provide formal, although linguistically implausible alternatives.

It is at this point that we need an interpretation theory relating the formal properties of the rivaling formalisms to the data: cohesion judgments.

CHAPTER 4 / ALTERNATIVE DETERMINISTIC MODELS FOR THE CONSTITUENT FORMALISM AND THE DEPENDENCY FORMALISM

4.1 INTRODUCTION

On Page 41 we mentioned Levelt's challenge, directed at those linguists who might be inclined to reject his interpretation of the C-grammar rather than the C-grammar itself, to present an alternative interpretation theory. In Section 2.4 arguments of various sorts have been put forward for rejecting both CGL and DGL and for postponing, at least temporarily, any decisions on the grammars themselves. Answers to the call for alternative interpretation theories will therefore form the subject matter of the present chapter. From our methodological point of view adopted in Chapter 1 and elaborated in some detail in Section 2.4 such alternative interpretation theories involve alternative options with regard to the function F and the arguments in the formula $R = F(G;\ e_1,\ \ldots,\ e_k,\ \mathcal{E})$. These options will be dealt with in the present and subsequent chapters.

In the present chapter we shall concentrate mainly on an alternative function F which will not of course be based on the notion of a distance function, but on that of "syntactic completeness", to be explained shortly. The formal properties of the syntactic structures here involved, the way these are hypothesized to determine cohesion and how this determination is conceived will all be specified. Initially in this chapter, these details will be summarized in two deterministic models, one for the C-grammar and one for the D-grammar, although we believe that ultimately (see subsequent chapters) probabilistic integrated models will prove indispensible for cohesion judgments.

This prior deterministic formulation has been deliberately chosen for two reasons. Firstly, even if probabilistic integrated models should eventually prove necessary, the ultimate goal of cohesion research remains the evaluation of the relative structural adequacies of the grammars involved. For these evaluation purposes a clear delineation of what is regarded as comprising the syntactic part of the interpretation theory must be given. The deterministic formulations are intended to specify just this delineation in a clear and independent fashion.

Secondly, this deterministic formulation will enable a fairer and directer comparison of the basic ideas underlying our and Levelt's interpretation theories. Whether *syntactic completeness* can or can not furnish a better basis than *the inclusion relation over constituents* for the prediction of cohesion judgments should not be decided on by opposing an extremely vulnerable formalization of the latter idea to a much less vulnerable probabilistic shaping of the former. Following the presentation of these deterministic models we shall conclude this chapter with an explorative empirical evaluation of their relative merits and demerits.

4.2 SOME FURTHER CONSIDERATIONS OF THE BASIC IDEA UNDERLYING CGL

Before presenting our alternative deterministic interpretation theory for the C-grammar it will be useful to recall the basic idea underlying Levelt's interpretation theory. This interpretation hypothesized that cohesion is predictable from the inclusion relation over constituents. Whenever the smallest common constituent of two words, say i and j, is included in the smallest common constituent of another pair of words, say k and l, then the relation between i and j is stronger than that between k and l. The elaboration of this idea as CGL has been handled in detail in Section 2.3.

A nearly equivalent reformulation of CGL's interpretation axiom in terms of the set-theoretical representation of syntactic structures, given in Chapter 3, would read:

(4.1) For all words i,j,k,l, in the sentence:
$SCC(i,j) \subseteq SCC(k,l) \leftrightarrow r(i,j) \geqslant r(k,l)$
(with $\geqslant$ reflexive, antisymmetric and transitive, cf. Page 31).

Let us consider this formulation in some detail. Since the constituents in a C-structure are only partially ordered by the inclusion relation, one of the following possibilities holds for every quadruple of words i,j,k,l:

(1) $SCC(i,j) \subset SCC(k,l) \leftrightarrow r(i,j) > r(k,l)$

(2) $SCC(k,l) \subset SCC(i,j) \leftrightarrow r(k,l) > r(i,j)$

(3) $SCC(i,j) = SCC(k,l) \leftrightarrow r(i,j) = r(k,l)$

(4) $SCC(i,j) \cap SCC(k,l) = \emptyset$, the interpretation axiom does not apply.

In the first two cases strict inclusion holds and consequently the application of the interpretation axiom results in strict inequalities. In case 3 the SCCs are identical and the application of the axiom yields an equality (by antisymmetry). In case 4, occurring whenever the SCCs are involved in the "is to the left of" relation, the axiom does not apply as there is no inclusion in either direction.

This elaboration of the basic underlying idea as given at the beginning of this section is extremely vulnerable. That severe restrictions become imposed on the cohesion judgments is indicated by the ultrametric inequality deducible from this interpretation theory (see Page 37).

General experience in testing CGL -see Levelt (1974c)- reveals a dramatic failure of the equality predictions in contrast with the relative success of the inequality predictions. According to Section 2.4, however, many of the equality predictions should be regarded as byproducts of the interpretation theory. Consequently, their violation is not acknowledged as critical counterevidence against the basic idea underlying the model. At the same time, however, we see that the basic idea is supported by the promising inequality predictions, a fact which should not be overlooked when attempting to improve the interpretation theory.

Reformulation in terms of strict inclusion. Let us therefore investigate the consequences of a reformulation of the interpretation axiom in terms of strict inclusion, thereby deriving inequalities only. This might be done by replacing in (4.1) the symbols $\subseteq$ and $\geqslant$ by $\subset$ and $>$, yielding:

(4.2) $SCC(i,j) \subset SCC(k,l) \Rightarrow r(i,j) > r(k,l)$

This formulation, however, (as was the case with (4.1)) has the disadvantage of entangling two steps that can better be made separately. The first step is the deduction of the cohesiveness order over the word pairs according to the inclusion relation over their SCCs. This cohesiveness order should, as was the case with Levelt's cohesion function α, be conceived of as theoretical notion. The second step involves the introduction of numerical values $r(i,j)$ in accordance with the result of step 1. For the first step the formalization

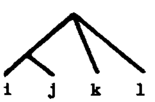
(4.3) $SCC(i,j) \subset SCC(k,l) \Leftrightarrow (i,j) \succ (k,l)$

would do (with $\succ$ the irreflexive, asymmetric and transitive relation "...is syntactically more cohesive than..."). The bidirectional arrow $\Leftrightarrow$ indicates that *syntactic* order relations can be derived if and only if the relevant SCCs are involved in a strict inclusion relation. Then the second step could be:

(4.4) $(i,j) \succ (k,l) \Rightarrow r(i,j) > r(k,l)$.

There would be less reason for considering these steps separately if the results of step 1 were to unambiguously determine the r-values on an ordinal scale. This, however, is not the case. Generally, the performance of step 1 -actually the syntactic basis of the predictions- results in a partial order. Consequently, there are some pairs of word pairs whose r-values are not assigned any ordinal constraints by the syntax. Accordingly, step 2 is formulated in (4.4) as a unidirectional implication. In order to obtain a completely determined ordinal scale, further constraints would have to be imposed. The primary concern of this section, however, is not scaling, but the evaluation of C-grammars. And the question we ask here is whether this evaluation can be based on the inequalities only, i.e. on the result of step 1. (4.3) has therefore been chosen as a less restrictive version of CGL, based on strict inclusion only, and for completeness the prediction has been added that whenever $(i,j) \succ (k,l)$, the empirical order will be $(i,j) > (k,l)$.

The central problem now becomes, whether the limited sets of restrictions imposed by (4.3) on the cohesiveness order for different C-structures are still sufficiently discriminative to justify the selection of a particular tree, or even to justify the choice of the C-formalism itself? Unfortunately, this is not generally the case. Certainly, many pairs of trees are assigned mutually contradictory inequalities by (4.3), so that critical observations could be made in order to choose between them. So would the PWC (i,j) versus (k,l) be a critical one for the choice between the C-structures $CS_1 = \{i,j,k,ij,ijk\}$ and $CS_2 = \{i,j,k,jk,ijk\}$, whose inequalities sets are $\{(i,j) > (i,k),\ (i,j) > (j,k)\}$ and $\{(j,k) > (i,j),\ (j,k) > (i,k)\}$ respectively. But for many other pairs of C-structures CS' and CS", viz. for all pairs of structures such that $CS' \subset CS''$, no such *contradictory* inequalities can be derived. Let us, for instance, compare the structures



and

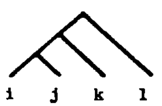


Figure 4.1

or, equivalently, $CS_1 = \{i,j,k,l,ij,ijkl\}$ and $CS_2 = \{i,j,k,l,ij,ijk,ijkl\}$ where $CS_1 \subset CS_2$. The inequalities set associated with CS_1 consists of the PWCs in which (i,j) is paired off with one of the other five "word" pairs. In all these cases (i,j) is predicted as the more cohesive pair. CS_2 gives rise to the same inequalities set plus the extra predictions that (i,k) and (j,k) are more cohesive than the pairs in which l is involved. Of course, these extra claims of CS_2 can in principle be tested, yet a selection between CS_1 and CS_2 cannot be coercively based on observations such as (i,k) $\gtrsim$ (i,l), since the observation is not incompatible with CS_1. Notice that this observation *would* have been incompatible with CS_1 in the (4.1) version of CGL, where an equality would have been predicted for this PWC.

The fact that (4.3) turns out to be insufficiently restrictive for making a choice between the structures CS' and CS" whenever CS' $\subset$ CS", carries further relevance for the problem of justifying the choice of the formalism itself. In Chapter 3 it was argued that C-structures and the constituency implied by D-structures differ in the respect of their *degrees*. D-structural constituencies are confined to degree 1, whereas C-structures are in the general case of higher degree. Consequently -as already mentioned in Chapter 3- much of the problem of choosing between the C- and the D-formalisms, will turn upon the question whether an empirically based choice can be made between constituencies of degree 1 or higher. These critical comparisons will generally concern structures CS' and CS" with CS' $\subset$ CS", as for instance is the case with $CS_1 = \{i,j,k,l,m,ij,lm,ijklm\}$ and $CS_2 = \{i,j,k,l,m,ij,lm,klm,ijklm\}$. For these comparisons the (4.3) formulation has too little discriminative power. Under (4.3) the claim that certain structures are essentially of degree 1 is indistinguishable from the claim that their degrees exceed 1. Where the highly vulnerable (4.1) version of CGL raises the likelihood of an error of the first kind, the (4.3) version tends to err on the other side.

A final consideration of both the (4.1) and (4.3) versions of CGL (which leads us as a matter of fact to our alternative interpretation theory) concerns the fact that both versions are stated in terms of *word* pairs only. The C-grammar, however, not only indicates what *words* combine to form constituents of a specified type, but also what *constituents* combine to form higher level constituents. In order to test these properties explicitly one would also have to collect cohesion judgments for pairs of constituents, following an extension of the domain of the interpretation theory so as to bring pairs of constituents under its scope. For the (4.3) version of CGL this might be accomplished by changing the opening words of the interpretation axiom: "For all words i,j,k,l, in the sentence..." into "For all constituents X,Y,Z,W, ($X \cap Y = \emptyset$ and $W \cap Z = \emptyset$) in the sentence SCC(X,Y) $\subset$ SCC(W,Z) $\Leftrightarrow$ (X,Y) $\succ$ (W,Z)". But this extension soon turns out to be empirically implausible in a very interesting way.

We shall illustrate the point by applying this "extended CGL" to the traditional C-structure for the sentence *the man hit a ball*. Using obvious abbreviations with the convention for simplifying the notation as introduced in Chapter 3, this structure can be expressed set-theoretically as CS = {T,M,H,A,B,TM,AB,HAB,TMHAB}. For the PWC (the man, hit a ball) versus (hit,a), thereby regarding words as one-element constituents, we would obtain the following deduction:

SCC(TM,HAB) = {T,M,H,A,B},
SCC (H,A) = {H,A,B},
SCC(H,A) $\subset$ SCC(TM,HAB) $\Leftrightarrow$ (hit,a) $\succ$ (the man, hit a ball).

Empirically, in all probability, the reverse judgment would be obtained.

So far, the problems associated with the idea that cohesion is determined by the inclusion relation over SCCs seemed to result from the way in which this idea was formalized. The relative success of the inequality predictions of CGL dissuaded us from rejecting the basic idea itself and encouraged us to bring the idea out of the range of the endocentricity argument by means of the (4.3) formalization, However, the above extension of CGL over pairs of constituents has led us to a crucial "observation" that seems to invalidate the very basic idea. Stated roughly, *relations high in a tree can be stronger than those low in a tree*. Or, more precisely, it is possible for certain building blocks of a sentence to cohere more strongly than other building blocks even when the SCC of the latter is included in that of the former. Syntactic relations with whole constituents can, and often will, be stronger than those with their subconstituents as, for instance, in *the boys went to Paris*, *went* and *to Paris* cohere more strongly than *went* and *to* or *went* and *Paris*.

In the next section we shall attempt to take this phenomenon into account by means of an alternative interpretation theory for the C-formalism. Given a C-tree, this interpretation will, on the level of word pairs, derive a nearly equivalent set of inequalities. It will avoid the equality predictions of the (4.1) version of CGL, thereby keeping the C-grammar out of the range of the endocentricity argument. Despite this, it will embody sufficient discriminative power to dinstinguish structures of degree 1 from those of degree > 1.

4.3 AN ALTERNATIVE DETERMINISTIC MODEL FOR THE CONSTITUENT FORMALISM

Let us reconsider the last two examples given in Section 4.2 where it was argued that the pair (went, to Paris) is intuited as more cohesive than (went,to) or (went,Paris) and (the man, hit a ball) as more cohesive than (hit,a). In both examples the former pair of constituents exclusively constitutes a higher level constituent, at least, if we argue in terms of the linguistically more usual C-structure. The pair (went, to Paris) depletes the verb phrase of the sentence; the pair (the man, hit a ball) depletes the whole sentence. It is in this sense that we could call these pairs syntactically complete: they exclusively constitute their SCCs. In the same sense, the latter pairs are incomplete: in (went,to) *Paris* is missing, in (went,Paris) *to* and in (hit,a) *ball*. The pair (the,to) in *the boys went to Paris* is even more incomplete, since the words *boys, went* and *Paris* would have to be added before their SCC becomes depleted. The pair (to,Paris), however, depletes its SCC and is syntactically complete. In our alternative interpretation theory for the C-grammar, henceforth CG2, we hypothesize that it is this notion of (in)completeness which determines cohesion. Syntactic relatedness is, in our view, not inversely related to the extension of the smallest common constituent, but to the incompleteness with respect to this SCC.

The formal elaboration of this point of view requires a definition of the notion *incompleteness* and, again, an interpretation axiom.

Definition (incompleteness)

The *incompleteness* of a constituent pair (X,Y), denoted ad I(X,Y), (with $X \cap Y = \emptyset$) is the set-theoretical complement of $X \cup Y$ with respect to the SCC(X,Y).

In other words, if we denote the union of two constituents X and Y as U(X,Y) then the following holds:

$$I(X,Y) = SCC(X,Y) - U(X,Y)$$

(where "-" stands for the set difference; A - B means $\{x \,/\, x \in A, \neg x \in B\}$

We shall illustrate this definition for the C-structure of Figure 4.2.

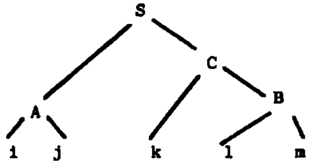


Figure 4.2

$I(i,j) = SCC(i,j) - U(i,j) = \{i,j\} - \{i,j\} = \emptyset$

$I(i,k) = SCC(i,k) - U(i,k) = \{i,j,k,l,m\} - \{i,k\} = \{j,l,m\}$

$I(i,B) = SCC(i,B) - U(i,B) = \{i,j,k,l,m\} - \{i,l,m\} = \{j,k\}$

$I(A,C) = SCC(A,C) - U(A,C) = \{i,j,k,l,m\} - \{i,j,k,l,m\} = \emptyset$.

Whenever for two pairs of constituents (X,Y) and (W,Z) it holds that $I(X,Y) \subset I(W,Z)$, then the former incompleteness is said to be smaller than the latter and the pair (X,Y) is said to be syntactically more complete than (W,Z). The interpretation axiom relates the syntactic relatedness order over the constituent pairs to this inclusion relation over the incompletenesses.

Interpretation axiom (CG2)

(4.4) For all constituents X,Y,W,Z, such that $X \cap Y = \emptyset$ and $W \cap Z = \emptyset$,

$$I(X,Y) \subseteq I(W,Z) \Leftrightarrow (X,Y) \succcurlyeq (W,Z)$$

(with $\succcurlyeq$ the reflexive, antisymmetric and transitive relation "is syntactically at least as cohesive as..."*

For completeness we add to (4.4) the stipulation that whenever $(X,Y) \succcurlyeq (W,Z)$ it will be predicted that $(X,Y) \geq (W,Z)$.

The incompletenesses are only partially ordered by the inclusion relation. Therefore, one of the following possibilities holds for every PWC (X,Y) versus (W,Z):

(1) $I(X,Y) \subset I(W,Z) \Leftrightarrow (X,Y) \succ (W,Z) \Rightarrow (X,Y) > (W,Z)$

(2) $I(W,Z) \subset I(X,Y) \Leftrightarrow (W,Z) \succ (X,Y) \Rightarrow (W,Z) > (X,Y)$

(3) $I(X,Y) = I(W,Z) \Leftrightarrow (X,Y) \approx (W,Z) \Rightarrow (X,Y) = (W,Z)$

(4) $I(X,Y) - I(W,Z) \neq \emptyset$ and $I(W,Z) - I(X,Y) \neq \emptyset$; in this case the

* Cf. Page 31: analogously to the specifications given there $\succcurlyeq$ is decomposable into syntactic strict inequality, $\succ$, and syntactic equivalence, $\approx$.

axiom is inapplicable since there is no inclusion between the incompletenesses.

All of these cases occur when (4.4) is applied to the structure of Figure 4.2. Case 1 (and mutatis mutandis case 2) in the PWC (k,l) versus (j,l): $I(k,l) = \{m\} \subset I(j,l) = \{i,k,m\}$. Case 3 in the PWC (i,j) vs. (l,m): $I(i,j) = I(l,m) = \emptyset$. Case 4 in the PWC (i,m) vs. (j,l) : $I(i,m) = \{j,k,l\}$ and $I(j,l) = \{i,k,m\}$ with $I(i,m) - I(j,l) = \{j,l\}$ and $I(j,l) - I(i,m) = \{i,m\}$.

Main properties of CG2. We shall review the main properties of CG2 by referring to the C-structure of Figure 4.3 and thereby investigate whether the four claims made in the concluding paragraph of Section 4.2 are valid.

Figure 4.3

(1) *Relations with whole constituents are stronger than relations with their constituting parts.* Let us consider the PWC (the man, hit a ball) versus (hit, a). Application of axiom (4.4) yields:

$$I(TM,HAB) = \emptyset$$
$$I(H,A) = \{B\}$$
$$I(TM,HAB) \subset I(H,A) \leftrightarrow (TM,HAB) \succ (H,A)$$

For this PWC we had already derived the reverse prediction by CGL. As a second example we consider the PWC (the man, a) vs. (the man, a ball). For this PWC the following holds:

$$I(TM,A) = \{H,B\}$$
$$I(TM,AB) = \{H\}$$
$$I(TM,AB) \subset I(TM,A) \leftrightarrow (TM,AB) \succ (TM,A)$$

In general, whenever two constituents X and Y, for which an inclusion holds, say $X \subset Y$, are compared with a third constituent Z, such that $Z \cap X = \emptyset$ and $Z \cap Y = \emptyset$, then CG2 predicts a stronger cohesion for (Z,Y) than for (Z,X). This can be easily verified. From $X \subset Y$ and $Z \cap Y = \emptyset$ it follows that $SCC(X,Z) = SCC(Y,Z)$ (a); let the complement of X with respect to Y be X'_Y: so, $Y = X \cup X'_Y$ (b).

According to the definition of incompleteness:

$$I(X,Z) = SCC(X,Z) - U(X,Z) = SCC(X,Z) - (X \cup Z)$$

$$I(Y,Z) = SCC(Y,Z) - U(Y,Z) = SCC(Y,Z) - (Y \cup Z).$$

Substituting (a) and (b) yields:

$$I(X,Z) = SCC(Y,Z) - (X \cup Z)$$

$$I(Y,Z) = SCC(Y,Z) - (X \cup X_Y' \cup Z)$$

so, $\quad I(Y,Z) \subset I(X,Z) \Leftrightarrow (Y,Z) \succ (X,Z).$

(2) *CG2 abstains from predictions whenever two constituents which are involved in an "is to the left of" relation are compared with a constituent outside their SCC.*

For example, application of (4.4) to the PWC (hit, a) versus (hit, ball) yields:

$$I(H,A) = \{B\}$$

$$I(H,B) = \{A\}$$

Neither $I(H,A) \subset I(H,B)$ nor $I(H,B) \subset I(H,A)$, so the rule is inapplicable.

Here we see that CG2 shares the advantage of the (4.3) version of CGL by keeping the C-grammar out of the range of the endocentricity argument. Here the (4.1) version would have deduced an equality. In general, whenever two constituents X and Y, such that $X \cap Y = \emptyset$ are compared with a third constituent Z such that $Z \not\subset SCC(X,Y)$, then CG2 abstains from deducing a cohesiveness order. This can be demonstrated since the specifications given imply that $SCC(X,Y) \subset SCC(X,Z) = SCC(Y,Z)$ and that the composition of SCC(X,Y) is:

$$\underbrace{\underbrace{(X \cup Y \cup I(X,Y))}_{SCC(X,Y)} \cup Z \cup R}_{SCC(X,Z) = SCC(Y,Z)} ,$$

where R represents the incompleteness, possibly empty, of SCC(X,Y) and Z. From this composition it is immediately obvious that $I(X,Z) = Y \cup I(X,Y) \cup R$ and $I(Y,Z) = X \cup I(X,Y) \cup R$. Irrespective of whether I(X,Y) or R are empty or not, it is not the case that $I(X,Z) \subset I(Y,Z)$ or $I(Y,Z) \subset I(X,Z)$ and consequently no syntactic cohesiveness order is derived.

(3) *On the level of the word pairs there is a considerable overlap between the inequalities sets implied by CGL and CG2, so that most of CGL's inequality predictions for word pairs are preserved under CG2.*

The inequality predictions for the structure of Figure 4.3 as implied by the two interpretations are summarized in the graphs of Figures 4.4 and 4.5. The nodes of these graphs correspond to the word pairs; the arrows to the cohesiveness order (from strong to weak). Under the word pairs in CGL's and CG2's graphs the SCCs and the incompletenesses are given respectively.

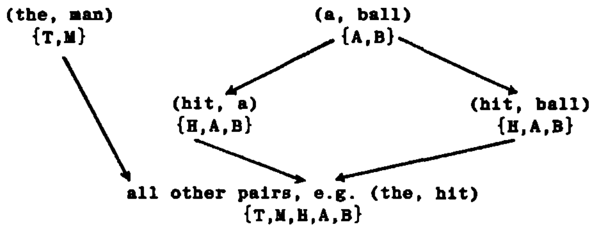


Figure 4.4 Cohesiveness order implied by CGL

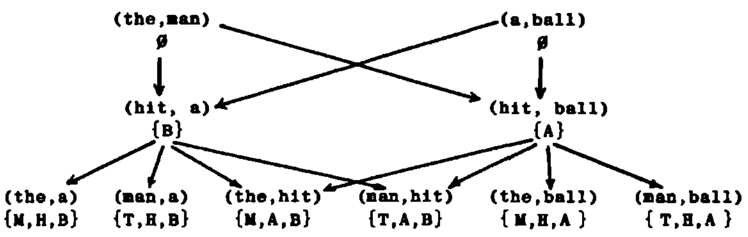


Figure 4.5 Cohesiveness order implied by CG2

Inspection of the graphs shows that in both interpretations (a,ball) is more cohesive than all other pairs except (the,man). In Figure 4.5 (the,man) dominates the same six pairs as in Figure 4.4 plus two extra pairs, viz. (hit,a) and (hit,ball) in an intuitively appealing way. In Figure 4.4 the pairs (hit,a) and (hit,ball) both dominate the same six pairs. In Figure 4.5 they each dominate 4 of these six pairs. There are no contradictions. The differences are confined to PWCs for which one of the interpretations shows an abstention as opposed to a prediction by the other.

Let us now, in general terms, investigate under what conditions CGL's inequality predictions for word pairs (i,j) versus (k,l) are preserved under CG2. We have to consider two possibilities; the pairs can be *disjoint* or *conjoint*. In the disjoint case neither i nor j equals k or l. In the conjoint case either i or j equals k or l. Of course, the case were both i = k and j = l is left out of consideration. In the *conjoint* case all CGL's inequalities are preserved under CG2. Suppose CGL derives the inequality $(i,j) \succ (j,k)$ which implies $SCC(i,j) \subset SCC(j,k)$. The composition of the SCC(j,k) then is:

$$\underbrace{\underbrace{\{i\} \cup \{j\} \cup I(i,j)}_{SCC(i,j)} \cup \{k\} \cup R}_{SCC(j,k)\ =\ SCC(i,k)}$$

where R represents the incompleteness of SCC(i,j) and k. From this it is easily seen that

$$I(j,k) = \{i\} \cup I(i,j) \cup R$$

This implies that $I(i,j) \subset I(j,k)$ and consequently $(i,j) \succ (j,k)$.

In the *disjoint* case, according to CGL, $(i,j) \succ (k,l)$ implies $SCC(i,j) \subset SCC(k,l)$. Let R denote the complement of SCC(i,j) with respect to SCC(k,l) so that:

$$SCC(k,l) = SCC(i,j) \cup R = \{i\} \cup \{j\} \cup I(i,j) \cup R.$$

Now we have to consider two subcases:

(a) both k and l are elements of R,

(b) either k or l is an element of R, the other being an element of I(i,j).

The case in which both k and l are elements of I(i,j) is excluded since it would contradict the strict inclusion $SCC(i,j) \subset SCC(k,l)$. In case (a) the composition of SCC(k,l) is:

$$SCC(k,l) = \{i\} \cup \{j\} \cup I(i,j) \cup \{k\} \cup \{l\} \cup S$$

with S the complement of $\{k,l\}$ with respect to R.

Hence, $I(k,l) = \{i\} \cup \{j\} \cup I(i,j) \cup S$ and again $I(i,j) \subset I(k,l)$ implies $(i,j) \succ (k,l)$ so that CG2 preserves CGL's prediction. The interested reader may verify this for the PWC (man,hit) versus (the,ball) where *the* and *ball* take the roles of i and j and *man* and *hit* the roles of k and l, I(i,j) is empty and S = {hit}.

In the disjoint case (b) CG2 abstains from predicting an order. Let us assume that $k \in I(i,j)$ and $l \in R$; let Y furthermore denote the complement of $\{k\}$ with respect to I(i,j) and U the complement of $\{l\}$ with respect to R. The composition of SCC(k,l) now becomes:

$$SCC(k,l) = \underbrace{\{i\} \cup \{j\} \cup \{k\} \cup T}_{SCC(i,j)} \cup \{l\} \cup U$$

Now it is easily seen that $I(i,j) = \{k\} \cup T$ and $I(k,l) = \{i\} \cup \{j\} \cup T \cup U$. Consequently, whether T and U are empty or not, $I(i,j) - I(k,l) = \{k\}$ and $I(k,l) - I(i,j) = \{i,j\}$ and the rule (4.4) does not apply.
This disjoint case (b) occurs, as can be verified in Figure 4.5, for the PWCs (hit,a) versus (the,ball), (hit,a) versus (man,ball), (hit,ball) versus (the,a) and (hit,ball) versus (man,a).

In concluding this consideration of the inequalities on the word pairs' level, we see that although some of the inequality predictions of CGL are not shared, they are also not contradicted by CG2. These differences are confined to the disjoint case (b). All other PWCs, both the conjoint and the disjoint case (a), induce the same predictions by CGL and CG2. Moreover, CG2 adds certain intuitively appealing inequalities for PWCs that fall outside the range of the (4.3) version of CGL, viz. all PWCs in which a complete word pair is paired with an incomplete one. The evaluation of the inequality predictions that are confined to either one of the interpretations is further a matter of empirical evidence.

(4) *In its capacity to discriminate different C-structures by deriving conflicting predictions for a number of PWCs, CG2 is intermediate between the (4.1) and (4.3) versions of CGL.* The (4.1) version allows for an unambiguous reconstruction of the C-structures on the basis of the set of derivable predictions, although at the price of an implausible ultrametricity assumption. In the (4.3) version -as we have seen- it is impossible to derive conflicting predictions for C-structures CS' and CS" whenever CS' $\subset$ CS", and thereby to rigorously distinguish between structures of degree 1 and higher. One Page 84 we emphasized the relevance of this type of discrimination for the central issues in this study: the comparison of the C- and D-formalisms. But CG2, with its express applicability to *constituents*, does embody this discriminative power to a large extent.

In describing the C-structure of a sentence, decisions will have to be made about whether or not to merge two constituents, say X and Y, into a higher level constituent. In terms of CG2, this will turn upon a choice between $I(X,Y) = \emptyset$ or $I(X,Y) \neq \emptyset$. The choice of $I(X,Y) = \emptyset$ will -apart from the inequalities of (X,Y) arising with all incomplete constituent pairs- produce as many equality predictions as there are other *complete* word and constituent pairs, (W,Z). CG2 would derive $I(X,Y) = I(W,Z) = \emptyset \leftrightarrow (X,Y) \approx (W,Z)$. But the choice of $I(X,Y) \neq \emptyset$ would result in $(X,Y) \prec (W,Z)$. So, if the decision is to assign the structure of Figure 4.3, with degree = 2, or that of Figure 4.6 with degree 1 to *the man hit a ball*

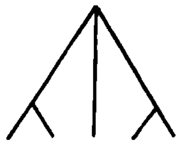


Figure 4.6

the PWCs (hit, a ball) versus (the,man); (hit, a ball) vs. (a,ball) and (hit, a ball) vs. (the man, hit a ball) are critical. In the structure of Figure 4.3 I(hit, a ball) = $\emptyset$ and for all PWCs equalities are predicted. In 4.6 I(hit, a ball) = {the, man} and inequalities are predicted.

It will appear that CG2 is closely related to the alternative deterministic interpretation theory for the D-grammar. The empirical evaluation of the former will therefore proceed more fruitfully when carried out in combination with that of the latter, so we shall postpone the presentation of the empirical evidence until Section 4.5. Our alternative dependency model comprises the subject matter of the next section. For the moment, this section on CG2 may be concluded with one final remark with respect to the "pureness" or "mixedness" of CG2, to use Levelt's terminology with which we became acquainted on Page 38.

From a theoretical point of view we adhere to Levelt's proposal that the linguistic component of an interpretation theory should comprise underlying syntactic structures rather than surface structures. In the 1974c version of this conception this led to the adoption of a transformational grammar as the linguistic component, with either a C- or D-grammar as its base. On the basis of the evidence available thus far (Levelt, 1974c; Fodor et al, 1980) this general point of view seems to be justified. Cohesion judgments for sentences with derived structures deviating substantially from the underlying deep structures have been shown to be more successfully relatable to the latter than to the former. But care must be taken not to overgeneralize these findings to other sentences types than those which have been studied in the studies referred to. Sentences might, and in fact will be encountered for which the cohesion judgments strongly suggest that surface structure is a co-determining factor. This co-determination, however, does not seem to be so strong that the surface structure should constitute the G in Formula 2.4; a better decision, perhaps, will be to account for possible surface structure effects by means of the e's in the same formula.

From a practical point of view, however, we shall hardly be concerned with this problem explicitly. Cohesion models are, for the time being, still in their very first phase of development. Of course, at this early stage many difficult problems can be pointed to, but even the simplest, such as giving a satisfactory account of very simple declarative sentences, have not yet been solved. It is for *this* reason -not because we believe that a full account of cohesion judgments can be given on the basis of deep structure relations only- that we shall confine ourselves to these simple declarative sentences. Such sentences already suffice to exhibit the properties that require either a D- or a C-grammar approach, as has been said in Chapter 3, where all the relevant issues could be explained for the sentence *the man hit a ball*. For these simple declarative sentences then, the question of the choice between pure or mixed models is of less relevance. We are still able to pursue the main purpose of this study even when bypassing this -in itself very important- question.

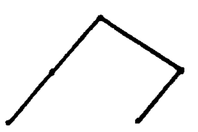
4.4 AN ALTERNATIVE DETERMINISTIC MODEL FOR THE DEPENDENCY FORMALISM

In considering DGL we must bear in mind a distinction similar to that in connection with CGL. On the one hand there is a *basic idea* about the relation between certain formal aspects of the D-structure for a given sentence and cohesion judgments; on the other hand there is the *formal elaboration* of the idea. As we have seen, the elaboration amounts to the definition of a cost distance over the connectedness graph associated with the D-structure, together with an interpretation axiom that inversely relates cohesion to distance. The arguments given in Section 2.4, however, have led to our abandonment of distance models for cohesion judgments. But how essential, we may ask ourselves, is the notion of a distance function for the derivation of the cohesiveness order that DGL specifies for the word pairs of a given sentence? Let us consider the partial cohesiveness order over the word pairs of the sentence *the pianist plays beautifully* that was given in Figure 2.9. It can be easily verified that this set of predictions is invariant under an alternative formulation of DGL specifying a word pair (i,j) as more cohesive than a pair (k,l) whenever the path connecting i and j is a subpath of that connecting k and l. For such a reformulation of the interpretation theory the exclusion of distance models would have no relevance.

Two extra requirements. In our opinion, however, there are other theoretical arguments for demanding more from an interpretation theory for a D-grammar than this "modified DGL" does. Firstly, as Levelt (1974c, p. 53 ff.) points out, the cohesiveness order specified by DGL for a given D-structure is only dependent on the connectedness graph associated with this D-structure. The order would be invariant under the choice of an alternative central element of the D-structure and thereby under an essentially different definition of the direction of the dependency relations. This means that DGL is insensitive to one of the most essential features of the D-formalism. In this section, on the contrary, we hope to demonstrate that these essential features can be inferred from the relatedness judgments.

Secondly, we shall require the interpretation theory for the D-grammar to reveal those basic features which, according to Chapter 3, distinguish D-structures from C-structures. It was argued there that D-structures can be conceived of as specifying a kind of constituency. This constituency, however, generally differs from that constituency specified by C-structures: the set of complete substructures is a constituency of degree 1 and the set of substructures constitutes a non-hierarchical constituency in which "constituents" may overlap.

The second requirement could be easily met by merely applying CG2 to the set of *complete substructures* or to the set of *substructures* associated with a given D-structure. To give an example we shall consider the D-structure of Figure 4.7.



the man hit a ball

T M H A B

Figure 4.7

The set of complete substructures is:

$$DS_1 = \{T,M,H,A,B,TM,AB,TMHAB\}.$$

The set of substructures is:

$$DS_2 = \{T,M,H,A,B,TM,AB,TMH,HAB,TMHAB\}.$$

In order to distinguish DS_1 from the linguistically usual C-structures CS_1,

$$CS_1 = \{T,M,H,A,B,TM,AB,HAB,TMHAB\}$$

the PWC (the,man) vs. (hit, a ball) would be a critical observation, since application of CG2 to DS_1 would give the following result:

$$I(T,M) = \emptyset \ ; \ I(H,AB) = \{T,M\}, \text{ hence:}$$

$$I(T,M) \subset I(H,AB) \Leftrightarrow (T,M) \succ (H,AB),$$

whereas its application to CS_1 would yield:

$$I(T,M) = \emptyset = I(H,AB) \Leftrightarrow (T,M) \approx (H,AB).$$

In order to distinguish DS_2 from CS_1 the PWC (the,man) vs. (the man, hit) would be a critical observation. Application of CG2 to DS_2 would yield:

$$I(T,M) = I(TM,H) = \emptyset \Leftrightarrow (T,M) \approx (TM,H),$$

whereas its application to CS_1 would yield:

$$I(T,M) = \emptyset \subset I(TM,H) = \{A,B\} \Leftrightarrow (T,M) \succ (TM,H).$$

In order to discern whether the set of *complete substructures* (DS_1) or the set of *substructures* (DS_2), if either, plays a role in the determination of cohesion judgments, critical PWCs would be (the, man) vs. (the man, hit), (the,man) vs. (hit, a ball), (the man, hit) vs. (a,ball) and (hit, a ball) vs. (a,ball), as can be easily verified.

So, in facing the second requirement, CG2 might be proposed as an interpretation theory for both the C- and the D-grammar. Strong theoretical support for this proposal can be found in the fact that it simultaneously seems to meet the first requirement. Given the connectedness graph associated with the D-structure of Figure 4.7, the choice of a different node as central, say the one labeled "ball" (see Figure 4.8), would induce a different set of *complete substructures* or a different set of *substructures**

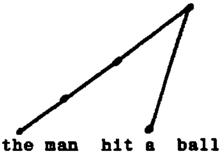


Figure 4.8

* The example is merely intended to illustrate one of the formal alternatives, irrespective of its linguistic implausibility.

These sets would be:

$$DS_3 = \{T,M,H,A,B,TM,TMH,TMHAB\} \text{ and:}$$

$$DS_4 = \{T,M,H,A,B,TM,AB,TMH,TMHB,TMHAB\}$$

respectively. In order to justify the choice of either *hit* or *ball* as central node of the D-structure, either DS_1 should be compared with DS_3 on the level of complete substructures or DS_2 with DS_4 on the level of substructures.

For the first comparison, among others, the PWC (the,man) vs. (a,ball) would be critical; for DS_1 this would yield:

$$I(T,M) = I(A,B) = \emptyset \Leftrightarrow (T,M) \approx (A,B),$$

for DS_3 the result would be:

$$I(T,M) = \emptyset \subset I(A,B) = \{T,M,H\} \Leftrightarrow (T,M) \succ (A,B).$$

For the second comparison a critical PWC would be (hit, a ball) vs. (a,ball), since for DS_2 CG2 would yield $I(H,AB) = I(A,B) = \emptyset \Leftrightarrow (H,AB) \approx (A,B)$ as opposed to $I(A,B) = \emptyset \subset I(H,AB) = \{T,M\} \Leftrightarrow (A,B) \succ (H,AB)$ for DS_4.

Of course, this discriminative power is very attractive from a theoretical point of view. However, whether this theoretically attractive property can be utilized remains an empirical question. An obvious requirement in this connection would be that the incompleteness idea should exhibit global empirical adequacy. It will appear that the most interesting competing structures for a sentence like *the man hit a ball*, for instance, DS_1, DS_2, CS_1 and $CS_2 = \{T,M,H,A,B,TH,AB,TMH,TMHAB\}$ have many predictions in common. The global empirical adequacy or inadequacy can be decided to a considerable extent on the basis of this great overlap in predictions. If CG2's basic idea, i.e. the incompleteness notion, should turn out to work on this overlap, then, as an additional step, the critical PWCs might be used to decide on the particular structure. But if it were found that this is not the case, the resort to critical PWCs would hardly be justifiable.

Before we turn to this important empirical question, we shall consider an extra theoretical and intuitively appealing possibility for enlarging the predictions set associated with a particular D-structure, by means of a more direct utilization of the dependency notion. As we have noted above, for many PWCs (X,Y) vs. (W,Z) CG2's interpretation axiom will be inapplicable. This happens whenever there is no inclusion relation between their incompletenesses. Among these PWCs are those cases where the members X and Y of an endocentric construction having, say, X as the head of Y, are paired with a constituent Z, that does not belong to the SCC(X,Y). As long as CG2 is applied to the C-formalism these abstentions are regarded as indicative of a natural interpretation, since the formalism itself does not specify the "head of" relation. The D-grammar, however, differs essentially from the C-grammar in that it *does* specify this relation. A natural requirement, therefore, would be to have an interpretation theory for the D-grammar utilize this information. To accomplish this we extend the incompleteness idea in a way which includes the afore-mentioned incompleteness principle as a special case, but adds to it the idea that a missing head renders a constituent pair more incomplete than a missing dependent. To enable this the interpretation axiom for the alternative D-model (henceforth: DG2) reads as follows:

Interpretation axiom (DG2)

For all PWCs of "constituents"* (X,Y) vs. (W,Z), such that $X \cap Y = \emptyset$ and $W \cap Z = \emptyset$:

$(X,Y) \succeq (W,Z) \leftrightarrow$ either (a) $I(X,Y) \subseteq I(W,Z)$
or (b) against every element i in
$I(X,Y) - I(W,Z)$ a unique element j in
$I(W,Z) - I(X,Y)$ can be set such that
j dominates i.

The predictions set resulting from application of DG2 to the set of complete substructures (henceforth DG2a) or to the set of substructures (DG2b) is the union of the predictions that would result from application of CG2 (condition a of the axiom) and those resulting from condition b. Two examples may illustrate the latter condition.

Example 1 concerns the cohesion of *hit* to the respective elements of the endocentric construction *a ball* in the set of complete substructures, DS_1, of the D-structure in Figure 4.7; in other words, it concerns the PWC (hit, a) vs. (hit,ball). $I(H,A) = \{T,M,B\}$; $I(H,B) = \{T,M,A\}$. $I(H,A) - I(H,B) = \{B\}$; $I(H,B) - I(H,A) = \{A\}$; B dominates A so, $(H,B) \succ (H,A)$.

Example 2 concerns the PWC (the,a) vs. (man,ball). $I(T,A) = \{M,H,B\}$; $I(M,B) = \{T,H,A\}$. For every element of $I(M,B) - I(T,A) = \{T,A\}$ a unique element in $I(T,A) - I(M,B) = \{M,B\}$ exists by which it is dominated: viz. *the* is dominated by *man* and *a* by *ball*. Hence, $(M,B) \succ (T,A)$.

* Actually, two variants of DG2 are involved: one in which "constituent" refers to *complete substructures* (as, for instance, in DS_1) and one in which *substructures* are referred to (as, for instance in DS_2).

Although the application of DG2 to DS_1 considerably increases the number of predictions over the number of CG2-predictions only, it does not deplete the complete set of PWCs. The PWC (hit,a) vs. (the,hit), for instance, amounts to a comparison of $I(H,A) = \{T,M,B\}$ and $I(T,H) = \{M,A,B\}$. The set differences in this case are $\{T\}$ and $\{A\}$. As *the* and *a* are not involved in a "dominates" relation, the incompleteness order is -without any further constraints- indeterminate.

4.5 *SOME PRELIMINARY EMPIRICAL RESULTS*

4.5.1 Introductory considerations

The interpretation theories introduced in the previous sections were formulated deterministically in terms of syntax only. In the first instance they specify a genotypic order relation $\succeq$ over the word pairs or the non-overlapping constituent pairs of a given C- or D-structure. This relation is reflexive, antisymmetric, transitive and weakly connected; it is to be read as "... is *syntactically* at least as cohesive as...". Application of the antisymmetry postulate leads to a further partitioning of $\succeq$ in the irreflexive, asymmetric and transitive relation $\succ$ (read: "...is *syntactically* more cohesive than...") and the reflexive, symmetric and transitive equivalence relation $\approx$ (read "...is *syntactically* as cohesive as..."). As a consequence, for all PWCs (X,Y) versus (W,Z), with $X \cap Y = \emptyset$ and $W \cap Z = \emptyset$, the following possibilities exist

1. $(X,Y) \succeq (W,Z) \wedge \neg (W,Z) \succeq (X,Y) \Leftrightarrow (X,Y) \succ (W,Z)$
2. $(W,Z) \succeq (X,Y) \wedge \neg (X,Y) \succeq (W,Z) \Leftrightarrow (W,Z) \succ (X,Y)$
3. $(X,Y) \succeq (W,Z) \wedge (W,Z) \succeq (X,Y) \Leftrightarrow (X,Y) \approx (W,Z)$
4. $\neg (X,Y) \succeq (W,Z) \wedge \neg (W,Z) \succeq (X,Y) \Leftrightarrow (X,Y)\ ?\ (W,Z)$

(where "?" means that the order is *syntactically* indeterminate).

For testing purposes, the models were then supplied with a rule of correspondence relating their deductive consequences $\succeq$ to predictions $\geqslant$ with regard to the empirical relation $\gtrsim$. Here, we deliberately distinguish notationally the syntactic order $\succeq$ from the predicted order $\geqslant$ in view of the fact that non-syntactic factors might prove to be indispensable parts of a more integrated model and could eventually cause the relations $\succeq$ and $\geqslant$ to differ in some cases. In CG2 and DG2, however, $\succeq$ and $\geqslant$ are isomorphic. For all relevant PWCs (X,Y) versus (W,Z) it is stipulated that:
$(X,Y) \succ (W,Z) \Leftrightarrow (X,Y) > (W,Z)$ and $(X,Y) \approx (W,Z) \Leftrightarrow (X,Y) = (W,Z)$.

On testing the predictions. One suggestion for the empirical evaluation of a particular C- or D-structure would be to present the relevant PWCs (i.e. those for which predictions are derived) to subjects taken from the speech community, asking them to indicate for each the more cohesive constituent pair or to give an equality judgment. Since the interpretation theories dealt with thus far have been formulated deterministically in terms of the ideal, i.e. syntactic case without an explicit account of random or systematic error, in principle no violations of the predictions would be tolerable and the subjects would have to respond in a uniform fashion. However, common practice with deterministic models is often different as can be seen from the following quotation taken from Runkel & McGrath (1972, p. 434).

> "One sometimes hears it said, or even sees it written, that deterministic models for analysis are not as useful as they might be because actual data so rarely fit the model; that is, the actual data rarely show the perfect pattern. But such an assertion shows poor understanding of the human spirit; humans are not often discouraged from using a good idea by a small discrepancy from reality. The fact is that deterministic models for data analysis have seen a good deal of use in recent years. Researchers manage this by using the principle of goodness of fit. If the data take a pattern that seems a sufficiently good approximation to the specifications of the model, they declare the model vindicated and the data as conforming to it - or at least those data that did indeed fall inside the specifications of the model".

In this vein one could decide not to demand uniformly given responses, but to accept the modal response as the datum for a given PWC, provided that its modality did not depart too much from uniformity.

The test for each PWC might then be tallied in one of the cells of a predictions by responses frequency table as in Figure 4.9. In order to obtain a global goodness of fit measure the proportion of violations could be calculated by dividing the sum of the off diagonal entries by the total number of tests made.

Taking these violation proportions as criteria, the relative structural adequacies of the competing C- and/or D-structures could be assessed, provided of course that the indices do not work out so high as to render the basic ideas of the interpretation theories implausible.

		responses:		
		$\lessdot$	$\doteq$	$\gtrdot$
	<			
predictions:	=			
	>			

Figure 4.9

Complications. Attention must, however, be drawn to a few complications arising in connection with this testing procedure. The first difficulty is a practical one and leads to a modification of the testing procedure. Other complications of a more fundamental nature arose during an early phase of a pilot study intended as an initial trial run of the above interpretation theories using the modified testing procedure. These complications will reveal the necessity of probabilistic and integrated interpretation theories in which stochasticity and non-syntactic factors become explicitly accounted for.

Numerosity of PWCs. The practical difficulty concerns the large number of relevant PWCs, even for such a simple declarative sentence as *the man hit a ball*.

Imagine we wish to evaluate the predictions derived by DG2b for the D-structure:

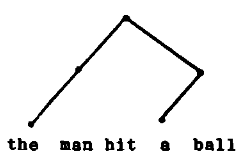


Figure 4.10

The relevant "constituents" in this case are: T,M,H,A,B,TM,AB,TMH,HAB and TMHAB.

These constituents combine into the 23 non-overlapping constituent pairs summarized and numbered in the cells of Table 4.1.

Table 4.1 Matrix of numbered non-overlapping constituent pairs for the structure of Figure 4.10

		1 T	2 M	3 H	4 A	5 B	6 TM	7 AB	8 HAB	9 TMH
1	T	-	1	2	3	4	-	5	6	-
2	M		-	7	8	9	-	10	11	-
3	H			-	12	13	14	15	-	-
4	A				-	16	17	-	-	18
5	B					-	19	-	-	20
6	TM						-	21	22	-
7	AB							-	-	23
8	HAB								-	-
9	TMH									-

In their turn the constituent pairs will combine to form $\binom{23}{2}$ = 253 dyads of constituent pairs. Among the 253 dyads there are in fact only 25 irrelevant PWCs for which DG2b derives no predictions, e.g. (man,hit) versus (hit,ball) and (the man, a) versus (the, a ball). Hence 228 PWCs would have to be presented in order to test all the consequences derived by DG2b for the above D-structure. Since the evaluation of the structural adequacy of a particular grammar would involve testing many other sentences of various types and lengths, it becomes clear that the above procedure would be impracticable.

The obvious alternative is to resort to one of the less laborious methods of data collection referred to in Chapter 2. The word sorting method is unsuited to the purpose of testing one of the main features of the incompleteness principle, viz. that relations to "wholes" can be stronger than relations to "parts". The reason is evident: it is impossible to sort two constituents into the same category without at the same time including the subconstituents and the words they consist of. Conversely, is possible to merge some words or subconstituents without respecting the integrity of a dominating constituent. The derived relatedness indices for lower level constituents therefore necessarily exceed those for higher level constituents.

Lastly, the method of rating scales would seem to be the preferable although not ideal alternative. The effort involved would become considerably reduced; for the sentence above only 23 constituent pairs would have to be presented.

Assessing empirical equality. But at the same time we are still plagued with the difficulty of inferring from the ratings one of the "empirical" judgments $>_e$, $=_e$ or $<_e$ for each PWC (X,Y) versus (W,Z). The inequalities are easier to decide on than the equalities. The inequalities might be inferred for those PWCs for which a statistical test of the difference between the mean ratings for (X,Y) and (W,Z) yields a significant result. But any temptation toward identifying the non-significant cases as equalities should be resisted, as this would amount to lumping together "genuine equality" with unreliable inequality. More appropriate to the deterministic formulation of the interpretation theories would, perhaps, be the procedure according to which we count the number of cases where (X,Y), compared with (W,Z) is assigned to a lower, equal of higher category, identifying the modal decision as the representative response, provided it does not depart too much from uniformity. But again, the equality category would be the problematic one. The number of equalities per subject is in part artificially determined by the number of categories on the scale. Other things being equal, it will be greater when a five point scale is used instead of a seven point scale. In the pilot study to be reported below, still another approach to handling or rather bypassing the equivalence problem has been chosen. It will be described in Section 5.4.2.

In any case, whichever of the above methods adopted, its greatest weakness will show up just where we need maximum reliability: in the assessment of empirical equivalence. For as we have seen in the antecedent sections, equality predictions play a diagnostic role in distinguishing the predictions sets associated with the competing structures. In this connection it is worth noting that the "equality-problem" would be much less of a headache in a probabilistic model specifying for each PWC (X,Y) versus (W,Z) the probabilities $p(XY.WZ) = p\,[(X,Y) >_e (W,Z)]$ and $p(WZ.XY) =$ $= 1 - p(XY.WZ)$. Let us assume that under two competing structural options the syntactic specifications for a given PWC were $(X,Y) \succ (W,Z)$ and $(X,Y) \approx (W,Z)$ respectively. Where the deterministic formulation would require a test procedure in which an empirical order or equivalence would have to be established, the probabilistic formulation would enable a choice to be made between two exactly specified alternative hypotheses, say p(XY.WZ) = 2/3 and p(XY.WZ) = 1/2, on the basis of the numbers of inequality judgments only. This choice might be based on a simple binomial evaluation of the observed choice frequencies.

A decision, however, to shape the basic syntactic ideas underlying CG2 or DG2 in a probabilistic fashion can not be exclusively based on theoretical considerations such as the one just put forward. Empirical reasons must be decisive here: the argument turns upon the character of the modal responses. We will have to contemplate the modal responses with two important questions in mind. Firstly, much will depend on the degree to which the modality of the responses departs from uniformity. Secondly, much will depend on whether in general the modal responses do fit into the pattern predicted by the model being tested. One ought to be prepared for two kinds of departure from the ideal case, in which uniformly given responses would conform to the model tested. (i) Were the modal responses, given either uniformly or not, unambiguously disconfirming, this would be a strong indication of systematic errors in the model. In this case, revision of the basic ideas underlying the model (e.g. the incompleteness principle) or of its basic assumptions (e.g. homogeneity of the speech community or the monopolistic role of syntax) should be preferred to probabilization. Under these circumstances, probabilization would only amount to the inoculation of a potentially invalid idea against counterevidence, thus leading to an unacceptable increase in the well known methodological risk of the second kind. (ii) Non-uniform, yet modally confirming responses would indicate that at the level of a single subject's choice behaviour errors occur, for which however, an interpretation in terms of random error might be acceptable. In this case, a probabilistic shaping of the basic ideas underlying CG2 or DG2 might be considered. If one is merely interested in modelling the modal behaviour of a speech community the deterministic formulation might be maintained. But if it is desired that the model should simulate choice behaviour at the level of a single subject, a probabilistic version will be preferable. Other considerations may also serve to tip the balance, e.g. the above-mentioned methodological advantage by which, under a probabilistic version, the problem of assessing empirical equivalence can be bypassed. Probabilization might then serve to protect a potentially valid idea against a too vulnerable formulation, thereby leading to a decrease in a risk of the first kind.

4.5.2 A pilot study

Bearing the above considerations in mind we shall now discuss some results of a pilot study, designed as an exploratory investigation into the interpretation theories dealt with thus far. Twenty five voluntary subjects (undergraduate students and members of the technical and administrative staff of the University of Nijmegen Psychology Department, both males and females, none of them linguists) took part in the experiment. Despite the weakness already recognized, the method of rating scales was adopted. The subjects were asked to rate all word and constituent pairs of several sentences of various types on a seven point scale. The Ss were run individually in experimental sessions of about 20 minutes each. In a written instruction the S's attention was directed to the fact that the words of a sentence cohere more or less strongly, depending on the way the sentence is constructed. The S was informed that he or she would be presented with certain Dutch sentences and then asked to express the degree of coherence of different pairs of words or word groups of these sentences on a seven point scale. An attempt was made to specify the notion of coherence in a syntactic sense. In doing so the word "syntax" was avoided but the possible interpretations of "coherence" were narrowed down by indicating, with examples, that: (i) coherence was not to be confused with mere superficial word distance, (ii) coherence was not to be confused with semantic relatedness, and (iii) coherence had to be judged only with reference to the sentence presented: accidental relationships existing between the words apart from the context of the sentence in question were to be regarded as irrelevant.

This study will not be discussed in full detail. We shall limit our attention here to the primary function of the results in the context of this study as a whole. Before discussing empirical details the main conclusions can be summarized in three points:

(i) The results suggested that the incompleteness principle represents a promising step forward. The violation proportions for some of the competing structures were quite low, and even then, many of the violations were far from definitive since the conflicting data was of moderate reliability.

(ii) It was not possible to evaluate the relative structural adequacies of the competing structural options on the basis of the violation proportions. Firstly, the aforementioned indeterminacy in delineating empirical equivalence implies consequences for the identification of violations which, in its turn, is indeterminate as well. Secondly, the violation proportions are based on totals differing to such a degree as to make them incomparable.

(iii) The data support a model in which stochasticity as well as determinants other than syntax should be incorporated.

We shall now illustrate these points with data obtained for one of the sentence types of the experiment, namely the one that has served as an example in the preceding sections and chapters: art_1 noun_1 $\text{verb}_{\text{transitive}}$ art_2 noun_2. Five versions of this sentence type were composed and each of these presented to five different Ss. The versions are: *de jongen slaat een vriend* (the boy hits a friend), *een meisje wast de auto* (a girl washes the car), *de vrouw vergeet een afspraak* (the woman forgets an appointment), *de man timmert het hek* (the man builds the fence) and *de leraar sloopt het hok* (the teacher pulls down the shed).

Table 4.2 Mean ratings given for the word and constituent pairs of the sentence type *de jongen slaat een vriend* (N = 25; SE = .28)

	J	S	E	V	DJ	EV	SEV	DJS
D	6.20	1.60	1.04	1.32	-	1.00	1.44	-
J		4.36	1.68	2.20	-	2.20	5.36	-
S			1.24	3.04	5.60	3.44	-	-
E				5.84	1.28	-	-	2.56
V					2.00	-	-	4.84
DJ						2.28	6.52	-
EV							-	5.68
SEV								-
DJS								

The data will be summarized in two ways. The first is in terms of the means of the 25 Ss' ratings assigned to each of the 23 word and constituent pairs. These means are given in Table 4.2 and represented on a scale in Figure 4.11. The initials of the first sentence above (D,J,S,E,V) will be used to denote the words and constituents. As an aid to the evaluation of the differences between means, the estimate of their standard error, $SE(\overline{X})$ = .282 (obtained by submitting the 23(PWCs) x 25(Ss) matrix of ratings to a one-factor repeated measurements-ANOVA), is displayed to the left on the scale.

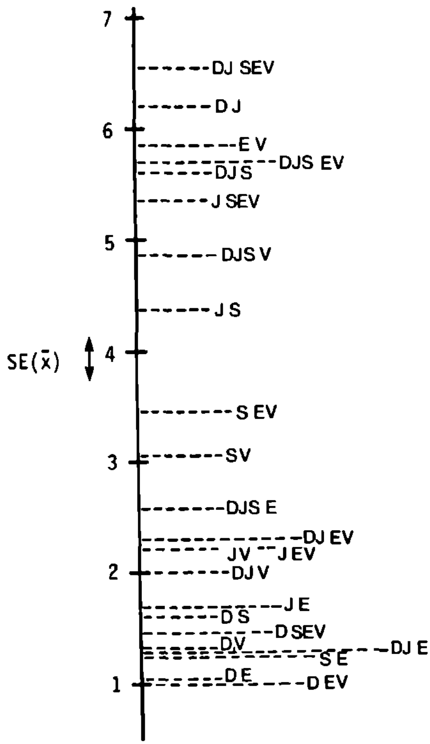


Figure 4 11 Word and constituent pairs of the sentence type *de jongen slaat een vriend* (DJSEV), scaled according to their mean ratings

The second summary, see Table 4 4, is in the form of a 23 x 23 matrix with rows and columns corresponding to the word and constituent pairs of the sentence, and with cells (i,j) indicating how many Ss rated the i'th pair higher than the j'th pair Cell (j,i) contains the number of Ss who gave the reverse judgment, whereas the sum of the entries in cell (i,j) and cell (j,i) subtracted from 25 yields the number of Ss that assigned an equal rating to the pairs i and j.

Table 4.4 Matrix of constituent pairs by constituent pairs; cells (i,j) contain the numbers of *Ss* who assigned the i'th pair to a higher category than the j'th pair.

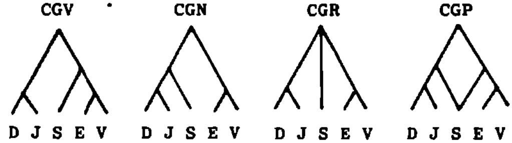
	D - J	D - S	D - E	D - V	D - EV	D - SEV	J - S	J - E	J - V	J - EV	J - SEV	S - E	S - V	S - DJ	S - EV	E - V	E - DJ	E - DJS	V - DJ	V - DJS	DJ - EV	DJ - SEV	EV - DJS
Nr	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0	22	22	22	22	22	19	21	22	22	17	22	21	17	19	2	22	19	22	22	22	7	9
2	0	0	8	7	8	3	0	5	4	2	0	6	1	1	4	0	6	2	3	0	3	0	0
3	0	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	1	1	2	0	2	1	1	2	1	1	1	2	1	1	2	1	2	1	1	0	1	1	1
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	4	9	8	9	0	1	8	5	3	0	6	1	0	3	1	7	3	5	0	4	0	1
7	2	20	22	20	22	21	0	20	20	21	5	21	14	2	13	4	22	16	21	8	19	1	3
8	0	4	6	6	6	4	2	0	3	3	2	5	2	1	2	0	5	5	3	2	3	1	1
9	1	9	13	12	13	10	1	10	0	4	1	11	4	1	4	2	11	7	7	0	6	1	1
10	0	12	15	15	15	12	1	11	6	0	0	13	4	0	4	1	13	7	8	0	5	0	1
11	4	24	24	23	24	24	14	22	22	22	0	24	21	5	16	5	24	18	23	12	23	1	5
12	0	3	3	3	3	2	1	3	2	1	0	0	2	0	0	0	3	0	2	0	2	0	0
13	0	15	18	18	18	16	4	17	15	15	2	16	0	2	7	2	17	13	15	1	14	0	0
14	4	24	25	24	25	24	17	23	23	25	10	25	22	0	20	5	24	19	25	17	24	2	5
15	1	14	16	14	16	14	6	15	15	15	4	16	11	3	0	3	16	11	15	6	15	0	3
16	0	21	21	21	21	21	17	19	20	20	16	21	19	16	18	0	21	18	21	20	20	6	9
17	0	1	5	5	5	3	0	3	1	1	0	5	0	0	1	0	0	2	2	0	2	0	0
18	0	11	10	11	11	10	4	9	8	8	2	10	5	3	6	1	10	0	8	4	6	0	0
19	1	9	12	12	12	11	1	8	7	4	0	10	3	0	3	2	12	6	0	0	4	0	0
20	3	23	24	23	24	23	11	22	20	22	4	23	19	5	14	5	23	18	24	0	23	1	4
21	1	13	15	14	16	13	0	11	10	7	1	14	3	1	5	2	15	7	8	0	0	0	2
22	4	25	25	24	25	25	18	23	23	25	14	25	24	12	21	5	25	22	25	20	25	0	7
23	2	21	22	22	22	22	15	20	20	21	11	22	20	10	19	3	22	18	21	16	21	2	0

The empirical adequacy of the incompleteness idea can, of course, only be assessed in application to a particular structure. Moreover, an index of success, for which we have proposed the violation proportion, will require a clear definition of empirical equivalence. This delineation could be based on establishing an interval, say Δ, such that whenever the mean for a constituent pair (X,Y) deviates less than Δ from the mean for constituent pair (W,Z), we say (X,Y) $=_o$ (W,Z). Under such a procedure evaluation of the incompleteness principle and its comparison with the "smallest common constituent principle" underlying Levelt's interpretation becomes very specific to the structure and delta chosen.

In order to obtain a less specific evaluation we therefore decided to apply both the incompleteness principle and the smallest common constituent principle to each of the four interesting syntactic options for the structure of this sentence (see Chapter 3). These options are {D,J,S,E,V,DJ,EV, SEV,DJSEV}, for which we choose the mnemonic denotation CGV, since it is characterized by the *V*erb phrase SEV; CGN = {D,J,S,E,V,DJ,EV,DJS,DJSEV} where the denotation CGN refers to the *N*ucleus DJS of subject and verb; CGR = {D,J,S,E,V,DJ,EV, DJSEV} where the R indicates a *r*educed constituent structure; and CGP = {D,J,S,E,V,DJ,EV,SEV,DJS,DJSEV} which is a non-hierarchical and therefore *p*seudo constituent structure because of the overlapping "constituents" DJS and SEV.

Violation proportions for different values of Δ, ranging in small steps from .0 to 1.6, were then calculated for each of the eight resulting combinations given in Table 4.5, together with their mnemonic labels.

Table 4.5 Conspectus of competing models



	CGV	CGN	CGR	CGP
smallest common constituent principle;	CGLV	CGLN	CGLR	CGLP
incompleteness principle	CG2V	CG2N	CG2R	CG2P

These violation proportions were then plotted as a function of the variable delta. The resulting graphs are given in Figure 4.12.

The curves clearly indicate what could have been predicted on a priori grounds, namely, that the smallest common constituent principle with its numerous equality predictions is favoured for large values of Δ since these yield many equivalences. The incompleteness models are by necessity dis-

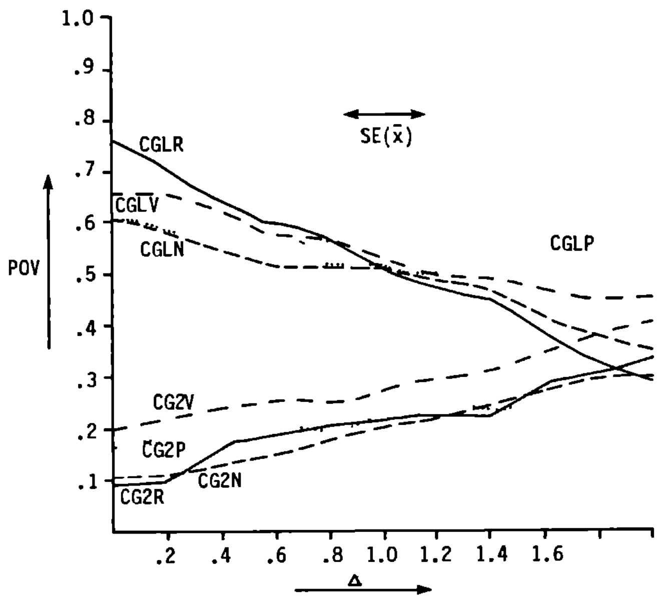


Figure 4.12 Proportions of violations (POV) for each of the combinations of Table 4.5, plotted as a function of different values of Δ.

favoured for large Δ's. An unacceptable increase in Δ to about 6 times the standard error of the mean would, however, be necessary to equalize both principles ' Conversely, small values of Δ inevitably disfavour the CGL versions without necessarily favouring the CG2 versions When, for a given PWC, a decrease in Δ results in re-interpreting an equality as an inequality, one of two possible changes takes place Either "genuine equivalence" changes into inequality by capitalizing on manifest random departures from equivalence, or "genuine inequality" becomes accepted. In the first case, it is very unlikely that a significant number of inequalities would fall into the pattern predicted by a particular theory This, however, is a fact which can be readily concluded from the POV-values for the incompleteness models at low levels of Δ.

This point can be illustrated by comparing the predictions vs. responses frequencies tables for CG2R at Δ = 1.6 and Δ = 0 in Tables 4.6a and 4.6b or those for CG2N at identical deltas in Table 4.6c and 4.6d.

Table 4.6 Predictions versus responses frequencies tables for CG2R and CG2N at different levels of Δ

(a)	$<_\bullet$	$=_o$	$>_\bullet$	Σ
<	17	15	0	32
=	0	1	0	1
>	0	0	18	18
Σ	17	16	18	51

CG2R: Δ = 1.6; POV = .294

(b)	$<_\bullet$	$=_o$	$>_\bullet$	Σ
<	28	1	3	32
=	0	0	1	1
>	0	0	18	18
Σ	28	1	31	51

CG2R: Δ = 0 ; POV = .098

(c)	$<_\bullet$	$=_o$	$>_\bullet$	Σ
<	43	20	0	63
=	1	4	1	6
>	2	8	29	39
Σ	46	32	30	108

CG2N: Δ = 1.6; POV = .296

(d)	$<_\bullet$	$=_o$	$>_\bullet$	Σ
<	58	5	0	63
=	2	1	3	6
>	2	0	37	39
Σ	62	6	40	108

CG2N: Δ = 0; POV = .111

From the above tables we see that the number of "empirical equivalences" decreases from 16 to 1 in the CG2R tables and from 32 to 6 in the CG2N tables. Most of these shifts from equivalence to inequality are in the direction of the predictions. This is the case for 11 of the 15 shifts in the CG2R tables and for 23 of the 26 shifts in the CG2N tables.

Comparing classes of models. Having, by the above approach, bypassed the problem of assessing empirical equivalence we may conclude that the comparison *between* both classes of models is in favour of the incompleteness principle. For the sake of completeness it is worth mentioning how the dependency models fit into the above picture. The POV graph for DGL would increase from .040 for Δ = 0 to .480 for Δ = 1.6. Those for DG2a and DG2b, in the same range, would increase from .080 to .396 and from .193 to .373 respectively. Hence if the POV graphs for the dependency models were to be plotted in Figure 4.12, they would fall more of less among the incompleteness graphs and share their general superiority to the CGL models.

Comparing models within classes. For more subtle comparisons *within* the class of those more successful incompleteness and dependency models this approach for bypassing the equivalence problem falls short. This is mainly because the data of this pilot study seems to be affected by a contaminating factor which the methodologist T.X. Barber (1973) denotes as "the failure to follow the protocol effect". In designing the experiment it was decided to present all word and constituent pairs (X,Y) in the order in which they appear in the sentence. This is a deviation from usual practice in presenting pairs of objects to Ss, namely, the randomization or counterbalancing of the order of objects over the presentations. The reason for this deviation lies in the fact that for some combinations the word order is indicative of the grammatical function. In these cases, reversal of the order would amount to a complication of the judgment process, since the subject would first have to mentally reconstruct the original order before giving his judgment. Without this mental reconstruction the judgment could be given without reference to its proper syntactic function. Reversal of the order of presentation might be expected to increase the probability of such a derailment in a way which differs for differing word pairs. Due to a misunderstanding in the execution of the experiment all word and constituent pairs were erroneously presented in the reverse sequence. Let us look at the consequence for our sentence *de jongen slaat een vriend*.

The isolation effect. For CG2V, for instance, the pair (S,EV), presented as (een vriend, slaat), is among the very critical ones. Its incompleteness is empty and consequently its cohesion should equal the cohesion of the pairs (D,J), (E,V), (DJ,SEV) and exceed all other cohesion values. The means, however, generally contradict this pattern. For (S,EV) we notice a mean of 3.44 as against 6.20, 5.64 and 6.52 for the above-mentioned pairs with which it should constitute an equivalence class. Furthermore, pairs like (S,J) with incompleteness {D,E,V} and (DJ,S) with incompleteness {E,V} should score lower than (S,EV), yet in fact yield means of 4.36 and 5.60 respectively. This corresponds with the graphs of Figure 4.12, where the models that take VP as a constituent (CG2V and CG2P) seem inferior to those that deny it (CG2N and CG2R). But just how valid are these means? The response pattern for the pair (S,EV) has some peculiar idiosyncracies with respect to most of the other response patterns. Its distribution, see Table 4.7, has a standard deviation SD = 2.20, which is relatively high when compared with the average value of the standard deviations of all pairs (1.36), and a clearly bimodal form.

Table 4.7 Response pattern for the pair (slaat, een vriend)

rating :	1	2	3	4	5	6	7
frequency:	9	2	1	2	6	3	1

Our *a posteriori* interpretation of this phenomenon ascribes it to the reversed presentation, which, apart from the sentence involved, could also be interpreted as a complete small sentence. A subject either complies with or forgets the instruction which asks him to relate his judgment to the sentence involved. If he forgets this, a relatively high judgment is to be expected since *een vriend slaat* can be conceived of as a complete isolated sentence when no reference is made to the experimental sample sentence. We will call this *the isolation effect*. When the subject does follow the instruction one of two reactions may occur. On the basis of the structure of the sentence involved the S realizes that *de vriend* is not the person doing the hitting as suggested by the presentation, but the person being hit. Where the S predominantly reacts in accordance with this deceptive suggestion low ratings are to be expected. Only when the S follows the letter of the instruction will he reconstruct the order *slaat - een vriend* and then judge this pair in a way which we hope will vindicate the adequacy of CG2V, CG2N, CG2R or CG2P. The fact that almost half of the subjects rated (S,EV) lower than the mean ratings for pairs such as (D,S), (D,E), (D,V), (D,EV), (J,EV), (J,E), (DJ,E) and (DJ,EV), strongly suggests that subjects who gave ratings of 1 or 2 only partially succeeded in obeying the instruction. While resisting the pitfalls of the suggestive sentence character of the presentation in which *een vriend* would be the agent, they failed to reconstruct the proper order and/or to relate it to the sentence involved. In our opinion the bimodal and heterogeneous response pattern for the pair (S,EV) -whether the above interpretation is correct or not- invalidates further reliance on its mean. Examination of the response patterns for certain other pairs point in the same direction. See, for instance, those for (J,S), presented as *slaat,jongen* and (S.V) presented as *vriend,slaat* (Table 4.8).

Table 4.8 Response patterns for the pairs (J,S) and (S,V)

rating :		1	2	3	4	5	6	7	$\overline{X}$	SD
frequencies:	(J,S)	3	0	7	1	6	5	3	4.36	1.87
	(S,V)	7	2	8	3	2	2	1	3.04	1.75

Since all these pairs are involved in PWCs critical to the comparative evaluation of the incompleteness models and of DG2 models, we must conclude that a true evaluation cannot be given on the basis of the pilot study here discussed.

4.5.3 Discussion

An obvious course would have been to replace the report of this pilot study with that of an improved replication. Nevertheless, we decided against suppression or replication for at least four reasons.

(i) Despite its weaknesses, the pilot study does illustrate that the incompleteness principle and the formalization of the dependency notion as given on Pages 87 and 99 are indeed sufficiently promising to deserve further study and development.

(ii) A replication of the seven point scale experiment, albeit with better controls for the "isolation effect" would offer no satisfactory solution to the crucial problem of testing the equality predictions.

(iii) The summary of the data, given in Tables 4.2 and 4.4 powerfully imply the stochastic or non-deterministic nature of the judging behaviour. Taken in combination with the reasons mentioned in (i) and (ii), this encourages a probabilistic elaboration of the basic ideas underlying CG2 and DG2.

(iv) The data of the pilot study are sufficient to indicate that cohesion is also determined by non-syntactic factors. In this study serendipity already revealed the isolation effect, with which we shall be explicitly concerned later. Some of these non-syntactic factors are a matter of good experimental practice and adequate controls, as in the effect just mentioned. But it is very likely that other determinants of intuited cohesion reflect inherent aspects of the process of making cohesion judgments which cannot be manipulated experimentally. Such factors would have to be explicitly accounted for in the model. The remainder of this chapter comprises a digression on this fourth point concerning non-syntactic determinants.

The problem of non-syntactic determinants. From the Tables 4.6a to 4.6d inclusive we see that the number of tests upon which the proportions of violations for the various models are based, differ to a large extent. The models delineate syntax differently so that the sets of PWCs for which they deduce predictions do not generally coincide. This, of course, complicates the methodology of evaluating these models. Were "ideal data" available syntactic predictions sets would suffice for deciding between particular C- or D-structures, or even between the formalisms themselves. "Ideal" here refers to what the data would look like if the internalized syntax were the only determinant of cohesion judgments. But this idealization, upon which is predicated the one-to-one relation between predictions sets and particular C- and D-structures is not yet justifiable. Apart from the isolation effect and certain other obviously non-syntactic factors such as those reported by Levelt (1969, 1970), there are occasions in which PWCs from outside the given syntactically relevant set, are judged in a very consistent way by the subjects.

We shall illustrate this point by means of some concrete examples taken from the experiment described. References to the results of the experiment will be to the information given in Table 4.4. Notation in the form "(X,Y) versus (W,Z): $n_1 - n_2 - n_3$", will mean that according to Table 4.4 n_1 subjects rated (X,Y) as more cohesive than (W,Z), n_3 subjects gave the reverse judgment and $n_2 = 25 - (n_1 + n_2)$ subjects assigned an equal category.

As we have already seen, the PWC (slaat,een) vs. (slaat,vriend): 2-7-16 is syntactically irrelevant for CG2V. This PWC, however, was presented to subjects because of its relevance to other models, and appears to be judged rather convincingly "in favour of" the pair (slaat,vriend). An advocate of DG2b may now reject CG2V's delineation of syntax and reinforce his position by stressing the fact that the modal response, (S,E) $\lessdot$ (S,V) corresponds to his predictions, since I(S,E) = {V} "governs" I(S,V) = {E}. A CG2V advocate will nevertheless be inclined to maintain his own delineation of syntax. For the non-syntactic explanation of the modal response, he might suggest that a factor like "semantic importance" is operative in those PWCs left unspecified by the syntax, as interpreted by CG2V. In his view, the judgment (S,E) $\lessdot$ (S,V) would be semantically based, since the missing word *vriend* contributes more to the meaning of the sentence than the missing word *een*.

Once the fact is acknowledged that non-syntactic factors may determine some syntactically *irrelevant* cohesion judgments, doubts arise as to whether the syntactically *relevant* PWCs are free from such factors. Reliance on the syntactically relevant PWCs would only be justified under the assumption that non-syntactic factors are only operative on PWCs left unspecified by the syntax. This is, however, demonstrably not the case. The proponent of CG2V may feel that he is not called upon to give an explanation for the PWC (S.E) vs. (S,V) as it does not belong to his predictions set. But the comparisons of either these pairs with the pair (S,EV) *are* relevant to his theory. The results for these PWCs are (S,EV) vs. (S,E) : 16-9-0 and (S,EV) vs. (S,V) : 11-7-7 respectively. The modal responses correspond to the predictions of the deterministic version of CG2V. But the difference in the "degree of modality" indicates that the first choice is made more decisively or with greater probability than the second. It would not be in accordance with CG2V to charge the syntax with the task of giving an account of this and similar observations. But the semantic explanation that was resorted to in order to account for the observation (S,E) vs. (S,V) : 2-7-16 might do well here too. In (S,EV) vs. (S,E), $I(S,EV) = \emptyset$ is compared with $I(S,E) = \{V\}$, whereas in (S,EV) versus (S,V) the comparison is between $\emptyset$ and $I(S,V) = \{E\}$, so that the first contrast is semantically much greater than the second. If, therefore, in dealing with PWCs relevant to CG2 models, we wish not only to specify order or equivalence, but differences in relative choice frequencies as well, incorporation of a factor such as semantic importance would seem to be indispensable.

This would constitute a nice example of the "realistic" approach to which we adhere. For the observables, in this case cohesion judgments, become accounted for by a model in which the syntactic and non-syntactic factors are necessarily dovetailed, and cannot be evaluated without taking this in consideration.

Similar remarks apply to DG2 models, for which a comparable point of view might be adopted. Again, weights might be associated with the elements of the D-structure, dependent on the degree of centrality in the D-structure. In this case the weights assigned should at least conform to the partial order of the dominance relation in the D-structure. Hence the result for the *syntactic* PWC (S,E) vs. (S,V) : 2-7-16 conforms with the prediction that DG2b derives for the set of substructures: $I(S,E) = \{V\}$ and $I(S,V) = \{E\}$ with V more central to the sentence than E. There are other PWCs, however, for instance (D,S) vs. (S,V) : 1-9-15, for which CG2V gives a syntactic account, since $I(S,V) =$ $= \{E\} \subset I(D,S) = \{J,E,V\}$ as opposed to DG2b where $I(D,S) = \{J\}$ is neither the head nor the dependent of $I(S,V) = \{E\}$. Challenged to give a non-syntactic explanation, the proponent of DG2b might, in his turn, resort to "semantic importance" or "analogy to the syntactic cases". Similarly, the advocate of DG2b cannot withdraw from a consideration of non-syntactic factors. He needs them to account for differences between relative choice frequencies in PWCs that he does regard as syntactic: (D,S) vs. (D,V) : 7-17-1 and (S,V) vs. (D,V) : 18-6-1, where $I(D,S) = \{J\}$, $I(S,V) = \{E\}$ and $I(D,V) = \{J,S,E\}$.

In summary, the results thus far reviewed strongly encourage us to abandon a deterministic formulation of the competing models and to reshape them in a probabilistic fashion. This could, moreovor, lead to a natural incorporation of non-syntactic factors. The suggestions given in the preceding pages with regard to the way this reformulation might proceed will be elaborated in the next chapter.

CHAPTER 5 / PROBABILISTIC MODELS FOR COHESION JUDGMENTS

5.1 THREE ALTERNATIVES FOR PROBABILIZATION

According to Restle (1971), the general problem underlying the construction of a choice theory lies in formulating how the various factors contributing to a probabilistic response are to be combined. In the same text Restle discusses three approaches to response probabilities, viz. the Thurstone model (1945), Luce's choice model (1959), in many texts referred to as the BTL or Bradley-Terry-Luce model, and his own set-theoretical model (Restle, 1961).

In the present study, one of the various factors going into a response is assumed to be syntactic structure, conceived of in the manner outlined in the previous chapter. In this chapter we shall present our attempts toward amalgamating each of the approaches discussed by Restle together with the syntactic incompleteness principle into probabilistic models for cohesion judgments. More precisely, we shall investigate whether the models in question are able to successfully predict the probabilities p(XY.WZ) of choosing constituent pair (X,Y) over pair (W,Z) as a function of the difference, the ratio or the set-theoretical compositions of the incompletenesses involved. As a prerequisite for the development of these probabilistic models first we must introduce a measure of the incompleteness of any particular constituent pair or of its component elements.

It will be remembered from the set-theoretical representation of sentence structure given in Chapter 3, that the symbol W was chosen to denote the set of word occurrences of a particular sentence s and W^* to denote the collection of all subsets of W. The elements, C, of W^* were called clusters. Let P be the set of non-negative real numbers $P = \{x \ / \ 0 \leqslant x\}$. We now define a *weight* function on the Cartesian product $W^* \times P$ which assigns to each $C \in W^*$, the "empty cluster" $\emptyset$ included, a number x, denoted by /C/. This weight function obeys the following postulates:

P1. $/C/ = 0$ if and only if $C = \emptyset$;

P2. if $C_i \subseteq C_j$ then $/C_i/ \leqslant /C_j/$, for any C_i, C_j ;

P3. $/C_i \cup C_j/ = /C_i/ + /C_j/$ if and only if $C_i \cap C_j = \emptyset$

Since a particular sentence structure is a subset of W^*, application of this weight function enables us to derive for all constituent pairs (X,Y) their weighted *union* /U(X,Y)/, their weighted smallest common constituent /SCC(X,Y)/ and their weighted incompleteness /I(X,Y)/. The interrelationship between these measures follows from the above postulates:

/I(X,Y)/ = /SCC(X,Y)/ - /U(X,Y)/.

An elaborated example for the sentence *the man hit a ball* (words again abbreviated to T,M,H,A,B) is given in Table 5.1, where purely for the sake of illustration, arbitrary weights 1, 5 and 6 have been assigned to the articles, nouns and the verb respectively.

Table 5.1 Illustration of the notion of "weighted incompleteness" worked out for the sentence *the man hit a ball*; syntactic option: CGV; weights of the words: 1,5,6,1,5 respectively

NO	PAIR	T	M	H	A	B	WEIGHTS	/I/
1	T,M	0	0	0	0	0	1	0
2	T,H	0	1	0	1	1	5	11
3	T,A	0	1	1	0	1	6	16
4	T,B	0	1	1	1	0	1	12
5	M,H	1	0	0	1	1	5	7
6	M,A	1	0	1	0	1		12
7	M,B	1	0	1	1	0		8
8	H,A	0	0	0	0	1		5
9	H,B	0	0	0	1	0		1
10	A,B	0	0	0	0	0		0

D . $\underline{1}$ = /I/

The weighted incompletenesses /I/ result from postmultiplication of a design matrix D with the column vector $\underline{1}$ of the afore-mentioned weights. D is a (10 x 5) matrix in which rows correspond to word pairs, columns to the "elementary constituents" T,M,H,A,B and cell entries indicate whether a particular row pair includes the column word involved in its incompleteness (cell value = 1) or not (cell value = 0); at least, as far as the syntactic option CGV (see Chapter 4) is concerned.

Application of the approaches to response probabilities discussed by Restle (1971) to these weighted incompletenesses result in the following models.

Model 1. The choice probabilities are inversely proportional to the weighted incompletenesses of the constituent pairs involved.

Interpretation axiom (under model 1):

For every dyad of constituent pairs (X,Y) and (W,Z), with $X \cap Y = \emptyset$ and $W \cap Z = \emptyset$, it holds that:

$$p(XY.WZ) = \begin{cases} \frac{1}{2} \text{ if } I(X,Y) = \emptyset \text{ and } I(W,Z) = \emptyset \ ; \\ \dfrac{/I(W,Z)/}{/I(X,Y)/ + /I(W,Z)/} \text{ in other cases.} \end{cases}$$

With some reservation regarding theoretical difficulties occurring whenever one of the incompletenesses is empty, this proposal can be characterized as an application of Luce's choice theory to the weighted incompletenesses. With empty incompletenesses 1 or 0 probabilities are derived, but unlike the standard procedure of eliminating the ever dominated alternatives from analysis, the word pairs involved are retained in the analysis since they inherently belong to the domain of the syntactic theory. Nevertheless, model 1 will henceforth be denoted as the *Luce-model*. However, the general characterization of choice probabilities being determined by the *ratio* of the weighted incompletenesses undergoes slight modification as it is formulated in the above interpretation axiom.

Examples.

1) Pairwise comparison (man,hit) versus (man,ball):

$$p(MH.MB) = \frac{/I(M,B)/}{/I(M,H)/ + /I(M,B)/} = 8\ /\ (7 + 8) = .533.$$

2) Pairwise comparison (the,man) versus (a,ball):

p(TM.AB) = .5 according to the first line of the interpretation axiom.

3) Pairwise comparison (the,man) versus (the,hit):

$$p(TM.TH) = \frac{/I(T,H)/}{/I(T,H)/ + /I(T,M)/} = \frac{11}{11 + 0} = 1.0.$$

Model 2. Not all elements of the incompletenesses determine the weights that are inversely proportional to the choice probabilities, but only the differential elements in the incompletenesses involved. A more precise formulation is the following interpretation axiom:

Interpretation axiom (model 2):

For every dyad of constituent pairs (X,Y) and (W,Z), with $X \cap Y = \emptyset$ and $W \cap Z = \emptyset$ it holds that:

$$p(XY.WZ) = \begin{cases} \tfrac{1}{2} \text{ whenever } I(X,Y) = I(W,Z) \; ; \\ \dfrac{/I(W,Z) - I(X,Y)/}{/I(W,Z) - I(X,Y)/ + /I(X,Y) - I(W,Z)/} \\ \text{in other cases, where } I(W,Z) - I(X,Y) \text{ is the set} \\ \text{difference of } I(W,Z) \text{ with respect to } I(X,Y). \end{cases}$$

With minor adjustments, this is the set-theoretical model developed by *Restle* and it will henceforth be referred to as such.

Again, some examples may serve as clarifying illustrations.

Examples.

1) Pairwise comparison (man,a) versus (man , ball),

The set theoretical composition of the incompletenesses of the alternative pairs are given in the Venn diagram of Figure 5.1. The diagram shows that $I(M,A) - I(M,B) = \{B\}$ and $I(M,B) - I(M,A) = \{A\}$.

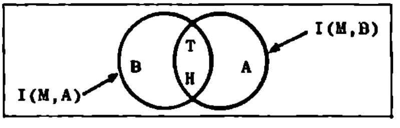


Figure 5.1 Venn diagram illustrating the relationship between I(M,A) and I(M,B)

Application of the second line in the interpretation axiom 2 results in:

$$p(MA.MB) = \frac{/I(M,B) - I(M,A)/}{/I(M,B) - I(M,A)/ + /I(M,A) - I(M,B)/} = \frac{/\{A\}/}{/\{A\}/ + /\{B\}/} = \frac{1}{1+5} = .167$$

whereas Luce's rule would have derived $8/(8 + 12) = .40$ for the same PWC.

2) Pairwise comparison (the,man) versus (a,ball):

$p(TM.AB) = \tfrac{1}{2}$ according to the first line in the interpretation axiom.

3) Pairwise comparison (hit,ball) versus (man,ball) :

$$p(HB.MB) = \frac{/I(M,B) - I(H,B)/}{/I(M,B) - I(H,B)/ + /I(H,B) - I(M,B)/} = \frac{7}{7+0} = 1.0.$$

Model 3. The choice probabilities are a normal ogive transformation of the differences between the weighted incompletenesses of the constituent pairs involved, these differences being expressed in terms of unit normal deviates.

Interpretation axiom (under model 3).
For every dyad of constituent pairs (X,Y) versus (W,Z), such that $X \cap Y = \emptyset$ and $W \cap Z = \emptyset$ it holds that:

$$p(XY.WZ) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z(XY.WZ)} \exp(-\tfrac{1}{2}z^2)\, dz$$

$$\text{with } z(XY.WZ) = /I(W,Z)/ - /I(X,Y)/.$$

Although this equation corresponds only to a particular case (viz. case V) of Thurstone's *law of comparative judgment* (see Torgerson, 1958, p.201), and is moreover adjusted to incorporate the incompleteness principle, model 3 will henceforth be referred to as the Thurstone approach. Its operation, again, will be illustrated by some examples. In this connection, obviously, we have to adjust the fictitious weights of the terminal elements used in this section to illustrate the probabilistic rules. Actually, the weights that constitute the weighted incompletenesses in the above formula -as is the case with models 1 and 2- have to be determined in the estimation phase of the data analysis. The details of this estimation procedure will be presented in Section 5.3. In the Thurstone approach, this estimation is subject to the condition that the normal ogive transformations of the differences of the resulting incompletenesses optimally match the empirical relative choice frequencies. Needless to say, the fictitious weighted incompletenesses in Table 5.1 give rise to differences that are entirely out of the range of a standard normal distribution. The obvious adjustment needed is to standardize the differences between the weighted incompletenesses of all $2 \times \binom{10}{2}$ ordered dyads of constituent pairs given in Table 5.1. These have a zero mean whereas their standard deviation equals 8.01. So in order to standardize these differences they have to be divided by 8.01. The weighted incompletenesses corresponding to these standardized differences as well, of course, as the weights of the elementary constituents, can be obtained by dividing the values given in Table 5.1 by this standard deviation. The rescaled "elementary weights" are .125, .624 and .749 for the articles, the nouns and the verb respectively, and the rescaled weighted incompletenesses /I(X,Y)/ are given in Table 5.2.

Examples.

1) Pairwise comparison (man,hit) versus (man,ball). In order to derive the probability, we have to subtract the /I(M,H)/ from /I(M,B)/ and to refer this normal deviate to the standard normal distribution. Hence, z(MH.MB) = = .998 - .874 = .124 with a cumulative probability of .549.

Table 5.2 Rescaled weighted incompletenesses I/(X,Y)/, corresponding to the standardized differences between the incompletenesses given in Table 5.1

	T	M	H	A	B
the		.000	1.373	1.997	1.498
man		-	.874	1.498	.998
hit			-	.624	.125
a				-	.000
ball					-

2) Pairwise comparison (the,man) versus (the,hit): z(TM.TH) = = /I(T.H)/ - /I(T.M)/ = 1.373 - .000 = 1.373 with a cumulative probability of .915.

3) Pairwise comparison (hit,ball) versus (man,ball): z(MB.HB) = = /I(M,B)/ - /I(H,B)/ = .998 - .125 = .873 with a cumulative probability of .809.

5.2 *THEORETICAL COMPARISON OF THE MODELS*

In the present section the distinctive aspects of the three probabilistic models introduced above will be demonstrated for the example given in Table 5.1. Data analytic problems regarding the estimation of the various weights, which have thus far been feigned arbitrarily, and the testing of the goodness of fit of the model, will be discussed in the next section.

For simplicity of notation the constituents of the sentence *the man hit a ball*, viz. {T},{M},{H},{A},{B},{T,M},{A,B},{H,A,B},{T,M,H,A,B} will be denoted without the usual braces and commas as T,M,H,A,B,TM,AB,HAB,TMHAB. The weights of the elementary constituents will be denoted by means of the corresponding lower cases, e.g. t = /{T}/, h = /{H}/ etc., consequently, /{T,M}/ = /TM/ = t + m and /{T,M,H,A,B}/ = /TMHAB/ = t + m + h + a + b.

Now let us suppose that two constituent pairs (X,Y) and (W,Z) are to be compared, with X,Y,W and Z varying over the above mentioned set of constituents and with both X ∩ Y and W ∩ Z empty. Then four types of PWCs can be distinguished according to the relation between the incompletenesses I(X,Y) and I(W,Z). These types are summarized in Table 5.3:

Table 5.3 Four types of PWCs according to the relations between the constituting pairs.

type	I(X,Y)	I(W,Z)	special features	example
1	$\emptyset$	$\emptyset$	-	(T,M) vs. (A,B)
2	$\emptyset$	not - $\emptyset$	-	(T,M) vs. (T,H)
3	not - $\emptyset$	not - $\emptyset$	$I(X,Y) \subset I(W,Z)$	(H,B) vs. (T,H)
4	not - $\emptyset$	not - $\emptyset$	no inclusion	(T,H) vs. (T,A)

Type 1 PWCs : both incompletenesses empty.
Example: PWC (the,man) versus (a,ball). If in this and all future cases we denote the probability of choosing the left hand pair over the right hand pair simply as p, then all three models derive in this case: $p = \frac{1}{2}$.
In the first two models this derivation follows by definition; in the Thurstone model it is because the normal deviate, corresponding to /I(A,B)/ - /I(T,M)/ equals zero.
Type 2 PWCs : only one of the incompletenesses is empty; example: (the,man) versus (the,hit), with $I(T,M) = \emptyset$ and I(T,H) = MAB. From I(X,Y) = $= \emptyset$ it follows that /I(X,Y)/ = 0; $I(W.Z) \neq \emptyset$ implies that /I(W,Z)/ is a positive number, say u. Moreover $I(X,Y) - I(W,Z) = \emptyset$, so /I(X,Y) - I(W,Z)/ = = 0 and I(W,Z) - I(X,Y) = I(W,Z), so /I(W,Z) - I(X,Y)/ = u. Therefore, both the Luce model and the Restle model result in: p = u : (u + 0) = 1.

According to the Thurstone model z(XY.WZ) = /I(W,Z) - I(X,Y)/ = = /I(W,Z)/ > 0. This implies: $\frac{1}{2}$ < p(XY.WZ) < 1. With the weights given in Table 5.1 and the rescaled weights given in Table 5.2 z(TM.TH) = 1.373 - 0 = = 1.373 so: p(TM.TH) = .915.

The conclusion is that in type 2 PWCs the Luce model and the Restle model derive deterministic predictions, whereas the Thurstone model derives a probability less than unity but greater than .5. The interesting empirical question, of course, is whether type 2 PWCs will elicit deterministic or probabilistic responses.
Type 3 PWCs: none of the incompletenesses is empty; one of them, say the incompleteness of the first pair (X,Y), is included in that of the other, (W.Z).

Let /I(X,Y)/ be u and let /I(W,Z)/ be v. Then, from $I(X,Y) \neq \emptyset$ and

$I(X,Y) \subset I(W,Z)$ it follows that $0 < u < v$. So, according to the Luce model, for the probability $p(XY.WZ) = v : (u + v)$ it must be the case that $\frac{1}{2} < p < 1$.

With respect to the Restle model we have to realize that $I(X,Y) \subset I(W,Z)$ implies $/I(X,Y) - I(W,Z)/ = 0$ whereas $/I(W,Z) - I(X,Y)/ = v - u$, which exceeds zero. The consequence is that under this model $p = (v - u) : (v - u + 0) = 1$.

As $/I(W,Z) - I(X,Y)/$ exceeds zero, so does $z(XY.WZ)$. Therefore, the probability under the Thurstone model again takes values between but not including .5 and unity. Again it has to be realized that (X,Y) is to be regarded as the left hand pair in the PWC.

Application of the successive models to the type 3 PWC (hit,ball) vs. (the,hit) yields:

$$p = \frac{/I(T,H)/}{/I(T,H)/ + /I(H,B)/} = \frac{(m + a + b)}{(m + a + b) + (a)} = \frac{11}{11 + 1} = .917$$

according to the Luce model;

$$p = \frac{/I(T,H) - I(H,B)/}{/I(T,H) - I(H,B)/ + /I(H,B) - I(T,H)/} = \frac{(m + b)}{(m + b) + 0} = 1$$

according to the Restle model; and

$$p(z < /I(T,H) - I(H,B)/) = p(z < 1.373 - .125) = p(z < 1.248) = .914$$

according to the Thurstone model.

The conclusion is that for type 3 PWCs the Restle model derives deterministic predictions, whereas the Luce and the Thurstone models derive probabilities less than unity but greater than .5. Again, the empirical question that should be asked is, whether type 3 PWCs give rise to probabilistic or deterministic response behaviour.

Type 4 PWCs: none of the incompletenesses is empty; moreover, there is no inclusion relation between them. These PWCs, actually the ones that are left unspecified by the deterministic version of the incompleteness models, are assigned probabilities between, but not including 0 and 1, by all three probabilistic models.

In summary, it can be concluded that the application of Restle's, Luce's and Thurstone's choice theories to the incompleteness principle results in an interesting sequence of models, two of which can be characterized as semi-probabilistic rather than probabilistic.

As far as the derivation of predictions is concerned, both the Restle and Luce approaches result in a partitioning of the set of PWCs into deterministic and probabilistic subsets. According to the Restle approach, the deterministic subset is made up of the type 2 and type 3 PWCs, viz. all PWCs with proper inclusion. Consequently, the delineation between the deterministic and probabilistic subsets exactly coincides with the delineation between syntactic and non-syntactic PWCs according to the deterministic version of CG2. The Restle model would require all syntactic PWCs to elicit deterministic choice behaviour. This, however, must be seriously doubted in view of the evidence discussed in the last few pages of Chapter 4. Although it is tempting to abandon the Restle approach for this a priori reason, we decided to suspend this decision until additional evidence has been gathered. One has to realize that the evidence in Chapter 4 stems from application of the seven point scale method and that things might turn out quite differently under the pair comparison method with its direct comparison of pairs.

According to the Luce approach, the deterministic subset reduces to the type 2 PWCs in which a syntactic complete pair is paired off with an incomplete pair. Deterministic choice behaviour for these PWCs will therefore be of critical importance to the choice between the Lucean and Thurstonian incompleteness models; the latter is entirely probabilistic.

The theoretical comparison concerning the partitioning of PWCs in probabilistic and deterministic subsets, thus far, is valid irrespective of the particular values that might be substituted for the parameters. Obviously we require more of a model of cohesion judgments than the correct specification of this partitioning: the resulting probabilities must also match the empirical relative frequencies. Of course, comparison of the competing models on this point presupposes a particular substitution of parameter values. The data analytic aspects of estimating these parameter values and testing the goodness of fit of the resulting representation will be handled in the next section. The estimation procedure, however, exhibits certain theoretical aspects as well. These will be attended to now.

The parameter space associated with the theoretical specifications given at the beginning of this chapter is characterized by a great latitude: all of the parameters t,m,h,a and b are allowed to range over the positive reals. This could turn out to be a hazardous point when attempting to compare the various models' goodnesses of fit. It is not unlikely that where a data analytic procedure is applied to an inadequate model, a poor fit may become obscured if the estimation procedure is allowed to "abuse" this latitude, by, for instance, manipulating the parameter values into an implausible constellation. An example may clarify this.

Let us suppose that, in reality, cohesion judgments follow Luce's rule, operating on CGR (see Page 110), with incompleteness weights $t = a$, $m = b$, and $t : m : h = 1 : 5 : 6$. For the PWC (man,hit) versus (hit,ball) this would imply:

$$p(MH.HB) = \frac{/I(H,B)/}{/I(H,B)/ + /I(M,H)/} = \frac{t + m + a}{t + m + a \; + \; t + a + b} = \frac{7}{7 + 7} = \tfrac{1}{2}.$$

Suppose further that in ignorance of the true state of affairs, we attempt a model in which we have Luce's rule operate on CGV. According to this model $/I(H,B)/ = a$ and $/I(M,H)/ = t + a + b$ so that, expressed in terms of the parameters:

$$p(MH.HB) = \frac{a}{a + (t + a + b)} = \frac{a}{2a + t + b}.$$

As far as this PWC is concerned, an estimation procedure designed in order to obtain an optimally fitting representation will have $a/(2a + t + b)$ approach $\tfrac{1}{2}$. This can be accomplished by reducing the parameters t and b as much as possible, i.e. as far as is compatible with the other PWCs' "dispositions" and with the model's specifications (e.g. positivity of weights). In the end we might find that the poor fit of an in itself inadequate model becomes reduced at the price of a very implausible asymmetry between the article parameters (e.g. $t << a$) or the noun parameters (e.g. $b << m$). A seeming fit would have been obtained instead of what should be preferred in this case: clear indications as to the inadequacy of the model and the origin of its inadequacy.

This potential degeneration of the parameter configuration, of course, reflects a shortcoming of a theoretical rather than a data analytic nature: the competing models are not restrictive enough. The obvious way out of this problem is to stipulate in advance a theoretical basis specifying which combination of parameters are to be excluded as implausible, together with the imposition of corresponding restrictions on the parameter space. Our proposal in this connection is to base these restrictions on two theoretical considerations: (i) words belonging to the same word category should receive equal weights; (ii) words belonging to the class of major syntactic categories should outweigh words belonging to the class of minor syntactic categories. For these restrictions little extra formal apparatus is needed beyond that given in Section 2.1. In both the C- and D-grammars the rules specifying lexical insertion can be conceived of as ordered pairs (X,x), where X denotes the lexical category to which the terminal element x belongs. Hence for every word pair (x,y) introduced by (X,x) and (Y,y) respectively, we simply stipulate that $X = Y \Rightarrow /x/ = /y/$. For the second restriction a bi-partitioning of the word categories into a subset V_{ma} of major categories (including, among others, verbs, nouns, abjectives and adverbs) and a subset V_{mi} of minor categories (including, among others, articles and prepositions) would have to be appended to the theoretical specifications so far given. For every word pair (x,y), introduced by (X,x) and (Y,y) respectively, we stipulate that $/x/ > /y/$ whenever $X \in V_{ma}$ and $Y \in V_{mi}$.

Application of these restrictions to the example given on Page 128 would result in assigning one and the same value, say p, to the articles *the* and *a*, a greater value, say $p+q$ (with q positive) to the nouns *man* and *ball* and another, say $p+r$ (with r positive), to the verb *hit*. Then, $/I(H,B)/ = /A/$ would equal p and $/I(M,H)/ = /TAB/$ would equal $p+p+(p+q) = 3p+q$. Hence, the expression for p(MH.HB), according to CGV-Luce, would read:

$$p(MH.HB) = \frac{p}{p+(3p+q)} = \frac{p}{4p+q} \quad .$$

A maximum of $\frac{1}{4}$ would be warranted, the possibility for arriving at a seeming fit would be reduced and the diagnostic function of the data analysis would be improved, i.e. its capacity to indicate *that* and *where* the model is deficient.

Weighted incompleteness versus distance. Finally, it is worth reminding the reader of the fact that in Chapter 2 distance models have been abandoned for cohesion. In this connection it is worth noticing that the weighted incompleteness model clearly does not belong to the class of distance models. To realize this we only need to think of a structure $\{i,j,k,ij,ijk\}$ with $/i/ \neq /j/$ so that $/I(i,j)/ = 0$, $/I(i,k)/ = /j/$ and $/I(j,k)/ = /i/$, which contradicts the triangle inequality.

5.3 DATA ANALYTIC ASPECTS OF THE PROBABILISTIC MODELS

For every PWC (X,Y) versus (W,Z), the data to be analyzed consists of the frequencies n(XY.WZ) and n(WZ.XY) of subjects giving the judgments $(X,Y) \succsim (W,Z)$ and $(W,Z) \succsim (X,Y)$ respectively. Two problems are involved in the analysis of these data: (i) estimating the values of the parameters; (ii) testing the goodness of fit of the resulting representation.

(i) *Estimating parameter values.* As for the determination of the "empirically optimal" configuration of parameters, we decide on the method of maximum likelihood. According to our models, any given values of the weights will permit calculation of the probability of the experimental data. According to the maximum likelihood criterion, the combination of weights for which this probability is maximized will be taken as optimal. Let us consider this in some more detail.

For every PWC (X,Y) versus (W,Z), the three probabilistic models presented above enable us to express the choice probabilities p(XY.WZ) as some function of the weights of the terminal elements of a sentence. In terms of our model sentence:

$$(5.1)\quad p(XY.WZ) = f(t,m,h,a,b)$$

The probability that among N subjects n(XY.WZ) will select (X,Y) as the most cohesive pair, and n(WZ.XY) select (W,Z), is -under the assumption that these choices can be regarded as mutually independent and of constant probability p(XY,WZ)- binomially expressible as:

$$(5.2)\quad p(n(XY.WZ)) = \binom{N}{n(XY.WZ)} p(XY.WZ)^{n(XY.WZ)}(1-p(XY.WZ))^{N-n(XY.WZ)}$$

The likelihood L of the observed experimental data D (i.e. the observed choice frequecies of all subjects for all PWCs) is then:

$$(5.3)\quad L(D) = \prod \binom{N}{n(XY.WZ)} p(XY.WZ)^{n(XY.WZ)}(1-p(XY.WZ))^{N-n(XY.WZ)^{*}}$$

The problem of finding that particular weight combination (t,m,h,a,b) which maximizes L(D) can be simplified by the consideration that the same combination of weights also maximizes any monotonic function of L, or equivalently, minimizes any monotonically decreasing function of L. The availability of James and Roos' (1974) minimization program MINUIT renders a reformulation of the problem of maximizing L in the form of a minimization of the monotonically decreasing function L* = - log L attractive:

$$(5.4)\quad L^{*} = -\sum \log \binom{N}{n(XY.WZ)} p(XY.WZ)^{n(XY.WZ)}(1-p(XY.WZ))^{N-n(XY.WZ)^{*}}$$

as alternative to the rather involved analytical procedure of taking partial

* The multiplication in (5.3) and the summation in (5.4) extend over all dyads of constituent pairs (X,Y) and (W,Z), such that $X \cap Y = \emptyset$, $W \cap Z = \emptyset$, and, as far as the Luce approach is concerned, $I(X,Y) \neq \emptyset$ and $I(W,Z) \neq \emptyset$. For the Restle approach the PWCs with $I(X,Y) - I(W,Z)$ or $I(W,Z) - I(X,Y) = \emptyset$ are excluded as well (see text).

derivatives of L with respect to all parameters and solving the equations:

$$(5.5)\quad \frac{\partial L}{\partial t} = 0,\ \frac{\partial L}{\partial m} = 0,\ \ldots\ ,\ \frac{\partial L}{\partial b} = 0.$$

The MINUIT program enables the application of a variety of minimization algorithms to a subroutine FCN, which is a user supplied subprogram containing the function to be analyzed, in our case function (5.4). The minimization algorithms included in MINUIT are James' Monte Carlo minimum search, Nelder and Mead's simplex method and Fletcher's variable metric method. All of these are utilized in sequence in our use of the program. We also make use of MINUIT's facility for influencing the progress of the algorithms, by fixing, restoring and setting limits to the variable parameters occurring in the function. This proves useful in handling the restrictions to be imposed on the model.

At this point a complication has to be considered. The above discussion of the probabilistic models has revealed that for some PWCs the Luce and Restle approaches result in deterministic predictions, i.e. in predictions of 1 and 0 probabilities. If the multiplication given in (5.3) were to extend itself over *all* dyads (X,Y) and (W,Z), then some factors in the product would equal zero, which, in turn, would render L(D)zero, regardless of the parameter values and the grammatical structures proposed. For this reason, the estimation procedure will be based only on those PWCs implied by the footnote on Page 130. This, of course, does not apply to the Thurstone model, as it is entirely probabilistic.

A final remark ought to be made about the estimation procedure and how the restrictions introduced in Section 5.2 are handled. In order to meet the first restriction, the word parameters are simply substituted by parameters for the word category to which they belong. In the case of our model sentence, for instance, an article parameter, say d, would have to be substituted for the word parameters a and t, a noun parameter, say n, for the word parameters m and b and a verb parameter v for the word parameter h. The resulting incompletenesses can be summarized in a 10 x 3 design matrix D^x (see Table 5.4) easily obtainable from the design matrix D, given in Table 5.1 and copied in Table 5.4. The first column of D^x, headed d, is the sum of the columns 1 and 4 of D, so that its cells indicate the number of articles in the incompleteness of the row pair involved. The second column of D^x is the sum of the columns 2 and 5 of D and indicates the number of nouns in the incompletenesses concerned. The third column of D^x simply copies column 3 of D and is the verb column, headed v.

The second restriction, to the effect that minor syntactic categories are not to outweigh the content words, can be expressed by writing the noun weight n as the sum of the article weight d and a surplus Δ_N, representing the difference n-d. Likewise, the verb weight will be expressed as $d + \Delta_V$ with $\Delta_V = v - d$. The incompletenesses can then be written as a function of d, Δ_N and Δ_V, and the second restriction is met by using MINUIT's facility for imposing lower limits (zero) on these parameters. It is easily verified that the resulting incompletenesses can now be summarized in a 10 x 3 design matrix D^{xx}, easily obtainable from D^x by copying its columns 2 and 3 and adding their sum to column 1 of D^x in order to form column 1 of D^{xx} (see Table 5.4).

Table 5.4 Recoding CGV's incompletenesses (D) according to the restrictions (1) t = a = d; m = b = n; h = v (D^x) and (2) Δ_N = n - d; and Δ_V = v - d (D^{xx}).

PAIR	D					D^x			D^{xx}		
	t	m	h	a	b	d	n	v	d	Δ_N	Δ_V
T , M	0	0	0	0	0	0	0	0	0	0	0
T , H	0	1	0	1	1	1	2	0	3	2	0
T , A	0	1	1	0	1	0	2	1	3	2	1
T , B	0	1	1	1	0	1	1	1	3	1	1
M , H	1	0	0	1	1	2	1	0	3	1	0
M , A	1	0	1	0	1	1	1	1	3	1	1
M , B	1	0	1	1	0	2	0	1	3	0	1
H , A	0	0	0	0	1	0	1	0	1	1	0
H , B	0	0	0	1	0	1	0	0	1	0	0
A , B	0	0	0	0	0	0	0	0	0	0	0

(ii) *Testing the goodness of fit.* Given the estimated parameters, for every PWC, the above-mentioned probabilistic models derive the theoretical choice probabilities and, thereby, the expected frequencies of subjects choosing one constituent pair over the other. An obvious criterion for an evaluation of the mismatch between the expected and empirically obtained frequencies would be the χ^2-test. The data of Chapter 4, however, suggest that rather extreme values (i.e. probabilities close to zero or one) can be expected for a large part of the PWCs.

The current χ^2-statistic might therefore be affected by the small expected frequencies artefact. It was thus decided to compare the expected and observed frequencies by means of the likelihood ratio approach (Mood & Graybill, 1963; Spitz, 1961) which is known to be less sensitive to extreme expected frequencies. We shall follow Spitz (1961) by denoting this test as the L-test. The formula, accomodated to our notation, is:

$$L = 2 \sum \left(n(XY.WZ) \ \ln \frac{n(XY.WZ)}{\hat{n}(XY.WZ)} + n(WZ.XY) \ \ln \frac{n(WZ.XY)}{\hat{n}(WZ.XY)} \right)$$

(see footnote on Page 130)

where $\hat{n}(XY.WZ)$ is the expected frequency of the choice $(X,Y) \succsim (W,Z)$ and the summation over those PWCs indicated in the footnote on Page 130. The statistic L follows the χ^2-distribution with df = $n_{PWC} - n_{par}$ (n_{PWC} is the number of PWCs *implied by the afore-mentioned footnote;* n_{par} *is the number of parameters* to be estimated).

In this connection, however, it should be remembered that the Luce and Restle approaches deduce deterministic predictions for some PWCs. These in turn involve zero expected frequencies and would make L incalculable. In view of this, as in the estimation procedure, the testing procedure partitions the set of PWCs into probabilistic and deterministic subsets. The probabilistic PWCs are then tested by means of the L-goodness of fit test; for the deterministic component, the mean absolute discrepancy between expected and observed relative frequencies, MAD*, is adopted as an ad hoc goodness of fit measure. Obviously, this partitioning must be kept in mind when determining the number of degrees of freedom needed for the evaluation of the L-value for the probabilistic subset of PWCs; a fortiori, any comparison of different models with different degrees of freedom will also have to take account of this factor.

* MAD-values, of course, are not necessarily confined to deterministic PWCs; overall MAD-values will prove useful in Chapters 6 and 7.

CHAPTER 6 / AN EMPIRICAL INVESTIGATION OF THE PROBABILISTIC MODELS

6.1 INTRODUCTORY REMARKS

This chapter covers the first empirical tests of the models that are implied by the various distinctions made in the previous chapters. Application of the probabilistic interpretations (Chapter 5) to the syntactic distinctions (Chapter 3) under a variety of restrictions on the parameter values, yields a large number of competing models. The testing of all these combinations with any reasonable set of basic sentence types would be an enormous enterprise. Although a definitive assessment of any given model's structural adequacy would demand testing with many different sentence types, in the present study we shall restrict ourselves to an investigation of those aspects essential to a clarification of the methodological issues involved. With such a plethora of competing models it is a good idea to eliminate the empirically less profitable as soon as possible.

Since every one of the competing models is, at least in principle, applicable to all sentences that can be generated by the particular grammar incorporated, every such sentence can be used as a test case for falsification purposes. In the initial stages of the empirical inquiry, therefore, much fruitful work can be done by means of a very few, or even a single sentence type. Another consideration in making such a restriction relates to the main goal of this study, which is the demonstration of the "realistic" methodological approach rather than an attempt at giving a definitive account of cohesion judgments. This demonstration of the realistic approach will also encounter no difficulties in the restriction to a single sentence type. For this purpose we shall return to that sentence type which has already served for illustrative purposes throughout this text, viz. art_1 $noun_1$ $verb$ art_2 $noun_2$, exemplified by *the man hit a ball*. It goes without saying that in our study with Dutch subjects, Dutch versions of this construction will be used.

Two experiments, henceforth to be referred to as Experiments 1 and 2, will be discussed in the following pages. Experiment 1 was a tentative pilot investigation which had to be called off after only a very few of the originally planned number of subjects had been tested. After some initial data had been gathered and submitted to the first orienting analyses, we discovered unexpected shortcomings in the design of the experiment. These shortcomings were likely to introduce considerable bias towards certain of the competing models and thus posed a serious threat to the validity of the experiment. In spite of its early cancellation Experiment 1 will be discussed in some detail. Many of the decisions reached with regard to overcoming problems in the design of the more successful Experiment 2 would remain puzzling without a review of the experience gained in this unsuccessful trial run.

6.2 EXPERIMENT 1

Problem. The empirical problems resulting from the theoretical issues covered in Chapters 3, 4 and 5 can be summarized in the condensed form of concrete questions. These questions pertain to the choice behaviour evinced by a sample of Dutch native speakers upon exposure to all dyads of word pairs from a sentence of the type art_1 $noun_1$ *verb* art_2 $noun_2$. More specifically, they concern the relative frequencies p(XY.WZ) and p(WZ.XY) of choices (X,Y) $\succ_o$ (W,Z) and vice versa, observed when these subjects are confronted with a pairwise comparison (X,Y) vs. (W,Z), where (X,Y) and (W,Z) range over all word pairs of the sentence *de man koopt het boek* (Eng.: the man buys the book). The questions are as follows:

Question 1. Are any of the ways in which the Luce, Restle or Thurstone approaches partition PWCs into deterministic and probabilistic subsets, as discussed in Section 5.2 and illustrated for convenience in Table 6.1, empirically justified?

Table 6.1 Matrix, indicating how the probabilistic models partition the PWCs into deterministic (D) and probabilistic (P) subsets, according to their types (Section 5.2)

	type 1	type 2	type 3	type 4
Restle	P	D	D	P
Luce	P	D	P	P
Thurstone	P	P	P	P

It goes without saying that a positive answer to this question would qualify as a necessary though not sufficient condition for the adequacy of the particular probabilistic approach.

It will prove useful to split question 1 into two subquestions.

Given a combination of a particular choice theory and a linguistic structure, (1a) do the PWCs that are specified as deterministic really elicit deterministic (or nearly deterministic) behaviour? (1b) Can an authentic application of the particular choice theory to the probabilistic PWCs, i.e. an application lacking any linguistic restrictions, fit the choice behaviour at all? It would be misguided to question the applicability of the choice theories under the severe restrictions of a syntactic model if even their unrestricted applications were to prove blatantly unsuccessful.

Let us consider the relevance of these questions for the different probabilistic approaches. As far as the *Thurstone* approach is concerned, question 1a does not apply. Table 6.1 indicates that all PWCs are probabilistic. This holds irrespective of the particular linguistic structure. Question 1b would then take the following form: is it possible to find scale values for the word pairs in such a way that the normal ogive transformation of their differences gives rise to choice probabilities that fit the observed relative frequencies in a statistically acceptable way? In this vein, question 1b qualifies as a preliminary test for further application of the Thurstone approach.

As far as the *Luce* approach is concerned, the deterministic PWCs are confined to type 2 PWCs in which one complete pair is paired off with an incomplete word pair. In all four linguistic options being tested, only (de, man) and (het,boek) are complete word pairs; the others are incomplete. This implies that these options agree in their delineation of the deterministic component. As a consequence, both question 1a -for the PWCs involving either (de,man) or (het,boek) as *one* of the pairs- and question 1b -for all other PWCs- qualify as a preliminary test for further application of the Luce approach. In this case, of course, question 1b asks whether it will be possible to assign scale values to the incomplete word pairs in such a way that these are inversely proportional to the choice frequencies in a statistically acceptable way.

The *Restle* approach makes all those PWCs deterministic in which the incompleteness of one word pair is strictly included in the incompleteness of another word pair (i.e. type 2 and type 3 PWCs). The competing syntactic structures differ with respect to this inclusion relation (see Chapter 4) and so the partitioning into deterministic and probabilistic subsets of PWCs varies with the syntactic structure involved. As a consequence, question 1 can not simply be posed once and uniformly for the various syntactic structures, as is the case with the other choice theories; it has to be raised anew for every syntactic structure. In this connection, however, the preliminary test will be confined to question 1a. As for question 1b, an authentic application of the Restle theory to the probabilistic PWCs would require the estimation of more parameters than there are PWCs (see Restle 1971). The Restle model only becomes interesting after additional restrictions have been imposed.

Question 2. How will the relative merits of the twelve models resulting from application of each of the three probabilistic interpretations to each of the four syntactic structures (CGV, CGN, CGR and CGP) turn out in terms of global goodness of fit, (a) when no extra restrictions are imposed on the parameter values beyond their positivity, (b) when the parameter values are to remain constant within word categories, i.e. /man/ = /boek/ = /noun/; /de/ = /het/ = /article/; (c) when it is additionally required that major syntactic categories should outweigh minor syntactic categories: i.e. /verb/, /noun/ > /article/?

Question 3. Where do the more promising combinations (if any) show their greatest deficiencies (again, if any) in terms of local contributions to the overall goodness of fit measures (L for the probabilistic PWCs; MAD for the deterministic ones; see Chapter 5)? What remedies are suggested by this diagnosis for improving the tentative version of the model?

Method

Material, design. Computer generated stimulus material was used for the presentation of all 45 PWCs constructible from the 10 word pairs of the sentence *de man koopt het boek*. Both the order of PWCs and of pairs within the PWCs were randomized afresh for every subject. The order of the words within the word pairs, however, conformed to their order in the sentence. These PWCs, preceded by training trials, were collected in small booklets of 50 pages, one page per PWC. Apart from the given PWC each page contained a replication of the entire sentence.

Procedure. The subjects were run individually in experimental sessions of about 10 minutes each. In a written instruction the S's attention was directed to the fact that the words of a sentence cohere more or less strongly, depending on the way the sentence is constructed. The subject was informed that he or she would be presented with a Dutch sentence and be asked to compare the degrees of coherence of different word pairs of that sentence.

An attempt was made to specify the notion of coherence in a syntactic sense. In doing so, the word "syntax" was avoided, but the possible interpretations of "coherence" were narrowed down by indicating, with examples, that:
(a) coherence was not to be confused with mere superficial word distance,
(b) coherence was not to be confused with semantic relatedness,
(c) coherence had to be judged with reference to the presented sentence; accidental relationships that might exist between the words apart from the context of the sentence in question were to be regarded as irrelevant. After reading the instruction the Ss had to work through the pages ot the booklet, underlining on each page in forced choice fashion that pair felt to be most cohesive.

Subjects. Twenty eight subjects, undergraduate students and members of the technical and administrative staff of the University of Nijmegen Psychology Department volunteered in the experiment.

Results

The choice frequencies obtained for the 28 subjects are shown in Table 6.2. Cell entries give the numbers of subjects that judge the corresponding row word pair as more cohesive than the corresponding column word pair.

Question 1 - Deterministic versus probabilistic PWCs.

As far as the Luce approach is concerned, both question 1a, pertaining to the deterministic PWCs, and question 1b, pertaining to the probabilistic

Table 6.2 Empirical choice frequencies for Experiment 1

	DM	DK	DH	DB	MK	MH	MB	KH	KB	HB
DM	-	28	28	28	22	28	25	26	25	21
DK	0	-	17	26	0	20	5	1	1	0
DH	0	11	-	21	0	12	2	2	0	0
DB	0	2	7	-	0	5	0	0	0	0
MK	6	28	28	28	-	28	26	20	21	4
MH	0	8	16	23	0	-	1	0	0	1
MB	3	23	26	28	2	27	-	15	3	2
KH	2	27	26	28	8	28	13	-	5	2
KB	3	27	28	28	7	28	25	23	-	1
HB	7	28	28	28	24	27	26	26	27	-

PWCs, are relevant. Stated more concretely, question 1a concerns the cells of the 1st and 10th row and column of Table 6.2 with the exception of cells (1,10) and (10,1) in which two rather than one complete word pairs are involved. Most of these cells do not contradict the expectation of deterministic or nearly deterministic behaviour, the mean absolute deviation (MAD) of observed relative frequencies and expected "probabilities" being .054. However, the PWCs (D,M) vs. (M,K) and (M,K) vs. (H,B) include non-negligible exceptions to which attention will be given in the discussion section.

The straightforward or linguistically unrestricted application of Luce's choice theory to the eight incomplete word pairs yielded the "unrelatedness" values given in Table 6.3. The values have been normalized such that their

Table 6.3 Unrelatedness values resulting from the linguistically unrestricted application of Luce's choice theory to the eight "incomplete" word pairs.

	D	M	K	H	B
D	-	(.000)	.099	.222	.944
M		-	.001	.222	.012
K			-	.008	.002
H				-	(.000)
B					-

raw sum of squares is unity.

An L-value of 20.56 was obtained, corresponding to a p-value of .486, the associated number of degrees of freedom being $\binom{8}{2} - (8-1) = 21$. The MAD-value amounts to .037. Keeping the small N of cases in mind, it can be cautiously concluded that the straightforward application of Luce's choice theory yields a promising result which encourages an investigation of its applicability under the restrictions of the linguistic models in this study. A final remark, however, should be made with respect to the single type 1 PWC in the data, lying outside the scope of question 1, viz. (D,M) vs. (H,B). Under all linguistic options, the choice percentages should be fifty-fifty, whereas the observed frequencies are 21 to 7, yielding an L-contribution of 7.32. In the discussion further attention will be devoted to this notable discrepancy between model and data.

Let us now consider the results for question 1b in connection with the Thurstone approach (in this case question 1a is not relevant). The linguistically unrestricted application of the Thurstone model yields the unrelatedness values given in Table 6.4.

Table 6.4 Unrelatedness values according to the Thurstone model

	D	M	K	H	B
D	-	.000	3.107	3.561	4.418
M		-	.841	3.547	1.930
K			-	1.741	1.192
H				-	.217
B					-

An overall L-value of 42.882 was obtained, interpretable as a χ^2 with an associated number of degrees of freedom of 45 - (10-2) = 37. The corresponding p-value is .234. The overall MAD is .040. Again a promising fit between model and data is found, which is not very surprising in view of the comparative literature on these choice theories. Differences in success, however, for the Luce and the Thurstone solutions could not be excluded in advance, since the preliminary test of the latter implies 17 additional PWCs, which are not involved in that of the former. Due to the paucity of data resulting from premature cancellation of the experiment we shall have to proceed with the analysis of both the Luce and Thurstone approaches; this is unavoidable in spite of the different assertions they make about the choice probabilities.

Questions 2a, 2b and 2c - Probabilistic models under different restrictions. This section discusses the application of the three probabilistic incompleteness models to the four syntactic options CGV, CGN, CGR and CGP under the restrictions that (a) word parameters should be positive, (b) word parameters should be constant within word classes, (c) parameters of words belonging to major syntactic categories should outweigh those of words belonging to minor syntactic categories. In the text, presentation of results will be confined only to those relevant to the main argument of this chapter. A more complete survey of results is given in Tables A1 to A7 inclusive in the Appendix.

Global inspection of these tables reveals that none of the twelve models is particularly successful. Even for the least unsuccessful models under the most moderate restriction (a) (see Tables A1, A2 and A3) the discrepancies between model and data are highly significant. This will appear from the following L-values whose associated p-values are all within an order of magnitude of 2×10^{-10}(!); CGP-Luce: L = 149.35, df = 25; CGP-Thurstone: L = 178.02, df = 42; CGR-Thurstone: L = 203.92, df = 42; CGP-Restle: L = 140.97, df = 17. Moreover, these analyses yield parameter values that are highly implausible from a linguistic point of view: very asymmetric weights were obtained for words belonging to one and the same word class. A not atypical example is the set of weights, obtained for the CGP-Luce model: /de/ = .017, /man/ = 1.000 (fixed in advance), /koopt/ = .316, /het/ = .028 and /boek/ = .134.

In the CGP-Thurstone model the verb parameter reaches the imposed lower limit of zero, thereby perhaps exhibiting a phenomenon comparable to the well known suppression effect in the context of multiple regression analysis.

On the one hand, the highly significant discrepancies between these models and the data hardly encourage further testing under the additional restrictions b and c. On the other hand, for diagnostic purposes it would not be advisable to neglect these analyses. A thorough diagnosis of the models' shortcomings could be handicapped by allowing the data analytic procedure to obscure a poor fit by organizing the parameter values into an implausible configuration. This was discussed at length in Chapter 5. As a matter of fact, some shifts in the various models' relative standings take place under the restrictions b and c.

This can be seen clearly from Figures 6.1 and 6.2 in which the models' relative merits are compared graphically. In these graphs the overall MAD values, i.e. the mean absolute deviations between empirical and expected probabilities for all 45 PWCs, are plotted for the various combinations of syntactic options (horizontal axis) and probabilistic rules (represented by separate lines). Figures 6.1 and 6.2 summarize the MAD-values under restrictions a and b respectively. No extra graph for restriction c is given, as nearly all representations obtained under restriction b already meet restriction c by carrying lower weights for minor categories. The only notable exception concerns the CGP-Thurstone model; its MAD-shifts under two different interpretations of restriction c are indicated by the small arrows labeled c1 and c2 in Figure 6.2. The reason for plotting MAD-values rather than, for instance, L-values is their direct comparability. Whereas the L-values are based on sets of PWCs varying both in composition and number, this is not the case for the overall MAD-values.

Figure 6.1 suggests what has already been mentioned in the text, viz. that the three CGP-models and the CGR-Thurstone model constitute the least unsuccessful quadruple of models. Figure 6.2 indicates that this is also the case under restriction b. However, the relative standings have changed. The CGR-Thurstone model rises from joint third place to joint first place. The MAD-values involved are as follows: CGR-Thurstone: .117; CGP-Thurstone: .117; CGP-Luce: .126; CGP-Restle: .125. As for the parameter values obtained under

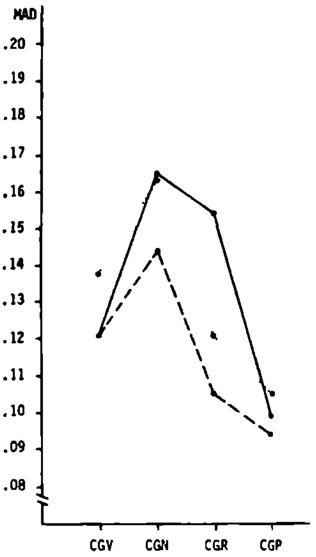


Figure 6.1

Experiment 1; overall MAD-values. No assumptions beyond positivity of weights (restriction a).

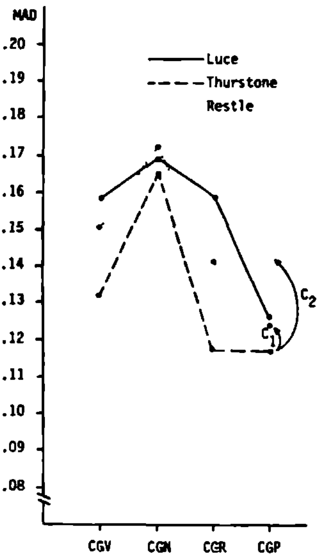


Figure 6.2

Experiment 1; overall MAD-values under restriction b: /de/ = /het/, /man/ = /boek/; for CGP-Thurstone, moreover, under restriction c1 and c2 (see text)

restriction b, in nine of the twelve models, viz. all CGV, CGN and CGR models, the noun parameters amply outweigh the article parameters and the verb parameters amply outweigh the noun parameters (see Tables A4, A5 and A6). A representative example is the CGR-Thurstone solution: /art/ = .049, /noun/ = .874 and /verb/ = 1.702. Hence for these models restriction c need not be tested separately since it has already been met by the solutions obtained under restriction b. The same applies to the CGP-Luce and the CGP-Restle models, whose respective parameter configurations are /art/ = .065, /noun/ = 1.000 (fixed in advance), /verb/ = .632 and /art/ = .184, /noun/ = 1.000 (again, fixed), /verb/ = .685. However, these two solutions differ from the nine referred to above in that the verb weight does not exceed the noun weight. In the CGP-Thurstone solution (still, under restriction b) the verb parameter even reaches the imposed zero lower limit and thereby lags behind the article parameter: /art/ = .879, /noun/ = 1.710 and /verb/ = .000.

Consequently the CGP-Thurstone model was studied under the restriction c. This was done in two analyses, one of them based on a weak, the other on a strong interpretation of this restriction (see Table A7). In the first analysis the parameter space was limited to the area where /verb/ ⩾ /art/. Since in the optimal configuration under restriction b the verb weight was less than the article weight, it was to be expected that this limitation would yield a solution with /verb/ = /art/. The second analysis was based on the requirement that, since it represents a major category, the weight of the verb should equal the noun rather than the article weight, and accordingly the limitation /verb/ ⩾ /noun/ was imposed on the parameter space. In the first analysis the resulting configuration was /art/ = /verb/ = .503, /noun/ = 1.381. The overall MAD increased from .117 to .124. The configuration obtained with the second analysis was: /art/ = .204, /noun/ = /verb/ = .976, yielding a MAD of .144. These two shifts are indicated by means of the arrows labeled c1 and c2 in Figure 6.2. For a fair comparison with the CGP-Luce and CGP-Restle models the effect of restriction c2, viz. /verb/ ⩾ /noun/, must also be reported (see Table A7). For CGP-Luce the following results obtain: /art/ = .073, /noun/ = /verb/ = 1.000, MAD = .127 (against .126 under restriction b). For CGP-Restle the results are /art/ = .186, /noun/ = /verb/ = 1.000 (fixed), MAD = .125 (as was the case under restriction b). These figures indicate that the /noun/⩽ /verb/ restriction hardly worsens the latter models' goodness of fit, whereas, on the contrary, it deprives the CGP-Thurstone model of its leading position among the least unsuccessful models. This illustrates nicely how our decisions with regard to the relative structural adequacies of the models are partly related to what we are willing to assume about the relative weights of the elements constituting the sentences in question.

At this point, however, a caveat should be mentioned. The small number of subjects employed (the consequence of certain design conditions to be discussed under question 3) ought to deter us from overrating the significance of the order of the least unsuccessful models. Moreover, the goodness of fit measures associated with these least unsuccesful models are so poor, that the only rational choice would seem to be to reject them. Whether this rejection should encourage adjustments of the models under consideration rather than an entirely different approach to cohesion judgments depends upon a careful diagnosis of the models' deficiencies. This diagnosis will be given in the subsection devoted to question 3.

Question 3 - sources of deficiencies.

The word pairs of the sample sentence can be partitioned into two subsets according to how their incompletenesses vary over the competing syntactic structures. One subset consists of six word pairs whose incompletenesses are constant over structures. These will accordingly be referred to as the "invariant" word pairs. Their incompletenesses are: $I(D,M) = I(H,B) = \emptyset$, $I(D,H) = \{M,K,B\}$, $I(D,B) = \{M,K,H\}$, $I(M,H) = \{D,K,B\}$ and $I(M,B) = \{D,K,H\}$. The remaining four word pairs -in which the verb is involved as one of the members- exhibit incompletenesses varying over structures as indicated in Table 6.5. These will be referred to as the "variant" word pairs.

Table 6.5 Incompletenesses of the "variant" word pairs

	(M,K)	(K,B)	(D,K)	(K,H)
CGV	D,H,B	H	M,H,B	B
CGN	D	D,M,H	M	D,M,B
CGR	D,H,B	D,M,H	M,H,B	D,M,B
CGP	D	H	M	B

Application of the restriction that weights should be constant over elements of the same word class leads to the weighted incompletenesses in Table 6.6, where we let *n* represent /M/ and /B/ and *a* represent /D/ and /H/.

Table 6.6 Weighted incompletenesses of the "variant" word pairs

	(M,K)	(K,B)	(D,K)	(K,H)
CGV	2a + n	a	a + 2n	n
CGN	a	2a + n	n	a + 2n
CGR	2a + n	2a + n	a + 2n	a + 2n
CGP	a	a	n	n

For the invariant word pairs, these weighted incompletenesses are /I(D,M)/ =

/I(H,B)/ = 0, /I(D,H)/ = 2n + v, /I(D,B)/ = /I(M,H)/ = a + n + v and /I(M,B)/ = 2a + v, where *v* denotes /K/.

The "variant" word pairs (and, of course, the PWCs in which they are involved) are of particular relevance for the choice between syntactic structures, whereas both these word pairs and the "invariant" ones are decisive for the adequacy of the general approach: the incompleteness principle in combination with one of the choice theories. In tracing the sources of deficiencies, we first concentrate on the variant word pairs, since syntax is our topic of primary interest; thereupon some of the "obstinate" invariant word pairs will receive attention.

Under the questions 2b and 2c it was found that the least unsuccessful models incorporate either CGR or CGP as syntactic structure. According to the above tables these structures can be characterized as *symmetric*. They differ from CGV and CGN in that the verb, rather than entering into a new constituent with a single constituent NP1 or NP2, it enters either with none of them (CGR) or with both (CGP). The tentative conclusion would be that, for the sentence in question, symmetry of structure is structurally more adequate than asymmetry. The high L-values, however, accompanying the least unsuccessful models force us to contemplate certain consequences of this conclusion.

Implications of symmetry of structure. Let us focus on the following four consequences that -among other things- are implied by symmetry of structure:

(a) p(MK.KB) = .5,

(b) for all word pairs (X,Y) other than (M,K) or (K,B): p(MK.XY) = p(KB.XY),

(c) p(DK.KH) = .5,

(d) for all word pairs (V,W) other than (D,K) or (K,H): p(DK.VW) = p(KH.VW).

A cursory look at Table 6.2 reveals that only requirement (b) is nicely met by the data: the relevant cells in the rows (or columns) headed (M,K) and (K,B), match quite well. Requirement (a) is violated by the frequencies obtained for the PWC (M,K) vs. (K,B), viz. n(MK.KB) = 21 and n(KB.MK) = 7. The quantity $|p(MK.KB) - \hat{p}(MK.KB)|$ = 7/28 = .25 amply exceeds the overall MAD-values (see Figures 6.1 and 6.2) for CGR-Thurstone, CGP-Luce and CGP-Restle. The contribution from this PWC to these models' overall L-values amounts to 7.32.

Under requirement c equal frequencies are expected for the PWC (D,K) vs. (K,H), whereas the observed frequencies are 1 to 27 ! This yields a contribution of 30.19 to the L-values of the models with symmetric structures. Suppose this PWC were to be considered in isolation; with the equality prediction dramatically overpredicting the intuited relationship of (D,K) and underpredicting that of (K,H), one might be inclined to incorporate CGV rather than a symmetric structure into the cohesion model. From the goodness of fit measures, however, as depicted in Figures 6.1 and 6.2, we already know that the adoption of the CGV-structure, although leading to local improvements, worsens the cohesion model globally. This neatly illustrates that isolated cohesion observations, as for instance Chomsky's observations quoted at the beginning of this study, carry no inferential force at all, unless the inference is supported by an elaborated interpretation theory. For a related assertion see Levelt (1974c, pp. 28-29).

Consequence (d), finally, requires the relevant cells in the rows headed DK and KH to equal each other pairwise. The frequencies for (K,H), however, all exceed those for (D,K), some of them quite convincingly.

The problem with the counterevidence reviewed is that it does not fit with any of the four patterns of incompletenesses in Table 6.6. Whereas the verb seems to cohere more strongly with the subject-N than with the object-N, there is an opposite cluster tendency as far as the articles are concerned. This seems to suggest that the prevalence of the symmetric structures over the asymmetric is a matter of their counterbalancing these opposite empirical relationships, rather than an indication of a genuine equality of the relationships of the verb to the first and the second NP.

In consideration of both the quantitative and qualitative aspects of the counterevidence it is easy to see that even the least unsuccessful of the twelve models must be rejected in their present form. More difficult is the decision as to what should or could be the next step in the approach toward cohesion judgments. Considerations with regard to future development of cohesion research will be given in the discussion.

Besides the afore-mentioned deficiencies with their direct relevance to the syntactic options in the cohesion model, one additional and more general source of violations is worth mentioning. For almost all PWCs in which the word pair (de,boek) occurs it is found that the predicted number of Ss choosing (de,boek) as the most cohesive pair far outweighs the observed number. This "overprediction" of the cohesion of (de,boek) occurs for all of the three least unsuccessful models and is of the same or even greater degree of magnitude as the "underprediction" of the pair (koopt,het). This appears clearly from the values of a statistic O(X,Y), especially designed to express the *degree of overprediction* of a particular word pair (X,Y). It is defined as the mean algebraic difference $\hat{p}(XY.WZ) - p(XY.WZ)$, with (W,Z) ranging over all other word pairs and with (X,Y) deliberately taken as the left hand term. A positive O(X,Y)-value indicates that (X,Y), on the average, is overpredicted by the particular model being tested, whereas a negative O(X,Y)-value indicates underprediction. For (de,boek) the O(D,B)-values obtained for the three least unsuccessful models, analyzed under restriction c, were: +.193 for CGP-Luce, +.160 for CGP-Restle and +.168 for CGR-Thurstone. For comparison, the respective O(K,H)-values, for (koopt,het) were: -.214, -.110 and -.104. Also, the probabilistic PWCs with (de,boek) as one of their members contribute disproportionately to the overall L-values. In the CGP-Luce model 7 PWCs with (de,boek) contribute a quantity 91.98 to the overall L-value of 229.43 (based on 29 PWCs). For CGP-Restle these figures are 96.40 (based on 5 PWCs) versus 207.85 (based on 21 PWCs); for CGR-Thurstone 86.44 (based on 9 PWCs) versus 253.80 (based on 45 PWCs). For the moment we restrict ourselves to signalizing this source of violations, postponing its discussion in terms of the "isolation effect" (see Chapter 4) to the next section.

Discussion

Our approach will now be to carefully examine the various aspects of the twelve competing models: two aspects which vary over the models and two which are constant. The varying aspects are (1) the choice of a particular *syntactic structure*, CGV, CGN, CGR or CGP; (2) the adoption of a particular *choice model*, Luce, Thurstone or Restle. The constant aspects, continuing the numbering, are (3) the adoption of the *incompleteness principle* and (4) the absence, thus far, of *non-syntactic determinants*.

As far as the *adoption of the syntactic structure* is concerned, we see no possibility of accounting for the conflicting evidence in a syntactic way. We maintain that if syntax is to be incorporated as one of the components of a model for the cohesion judgments involved, it should take one of the forms chosen. In other words, we persist in the opinion that the verb either clusters with NP1 (CGN), or with NP2 (CGV) or with both (CGP) or with neither (CGR), and that this exhausts the interesting possibilities.

As far as the *choice models* are concerned, at the moment we can see no compelling reason for eliminating one of them from further consideration. The authentic applications of the Luce and the Thurstone approaches were quite promising. The Restle approach could not be applied authentically, but its least unsuccessful restricted application, the CGP-Restle model, only slightly falls behind the CGR-Thurstone and the CGP-Luce model (see Figure 6.2).

These considerations make us incline to blame one of the remaining aspects of the interpretation theories for the dramatic increases in the L- and MAD-values that can be observed when the probabilistic models are applied under linguistic and interpretation theoretical restrictions. In other words, either the adoption of the incompleteness principle or the neglect of non-syntactic factors, if not both, should be questioned.

With respect to the latter aspect, it would be surprising rather than satisfying were the data in this experiment not to be affected to some degree by non-syntactic determinants. This was the case with the data of the experiment discussed in Chapter 4, and with almost all data reported thus far in experimentation on cohesion. There are indeed many instances in Experiment 1 where the local discrepancies between models and data strongly suggest alternative interpretations in terms of semantic factors, isolation effects, superficial word distance and the like. With such factors operating it is difficult to evaluate the third aspect of the interpretation theory, viz. the *incompleteness principle*, and especially to assess the type 1 and type 2 risks attendant upon a decision for rejecting or accepting it.

An attempt to trace the *non-syntactic factors* is thus imperative. The results of such an attempt could lead either to improved experimentation in which non-syntactic factors are avoided, or, were this to prove impossible, to the explicit incorporation of such factors into the cohesion model.

Bearing the above considerations in mind, we now raise the question of whether in the absence of a syntactic account a non-syntactic interpretation can be given for the conflicting evidence against symmetry of structure. More specifically, since there doesn't seem to be any chance of giving a syntactic explanation of *both* the greater cohesion of (man,koopt) in comparison with (koopt,boek) *and* the lesser cohesion of (de,koopt) in comparison with (koopt, het), we ask whether one or both of these conflicting tendencies can be accounted for in a non-syntactic way. Were plausible alternative interpretations to be found for both tendencies, a cohesion model with symmetry of syntactic structure might then still prove fruitful. We should, however, be prepared to consider the rehabilitation of asymmetry of structure, if reasonable interpretations could be found for only *one* of the tendencies.

Superficial word distance. One initial non-syntactic interpretation of the conflicting evidence is in terms of *superficial word distance*. Although experience with cohesion data does not suggest an important role for word distance, it is a factor that might become relevant under special circumstances. When subjects have to decide on one of the word pairs of an indeterminate PWC in forced choice fashion, there is an increased tendency to use criteria which would usually not be considered. This has also recently been observed by Fodor, Garrett, Walker and Parkes (1980, p. 366) in their employment of cohesion judgments. In this vein, subjects presented with the syntactically indeterminate PWC (man,koopt) vs. (koopt,boek) may predominantly choose (man,koopt) because *man* immediately precedes *koopt* whereas *koopt* is separated from *boek* by one word. For the same reason word distance might have caused (koopt,het) to be preferred to (de,koopt). A problem with this common interpretation for both tendencies is that the asymmetry of the frequencies involved is more extreme for the latter PWC (1 : 27) than it is for the former (21 : 7). Separate causes of the tendencies can not be excluded and will now be considered.

Isolation effect. A separate alternative factor that may in all probability have affected the PWC (de,koopt) vs. (koopt,het), either alone or in combination with superficial word distance, is the *isolation effect*. The term was introduced in Chapter 4 to refer to situations in which the combination of two words as a pair in a PWC suggests a relation or function differing from that holding between them in the sentence itself. Such an effect occurs when the subject reacts to the relationship suggested rather than the relationship *in the sentence*. The above-mentioned PWC contains (koopt,het) as one of its pairs. In Dutch, *het* is the written form of either the neutral definite article or the neutral pronoun (Engl.: *it*). The extremely asymmetric frequencies perhaps indicate that the subjects predominantly perceived the isolated word pair as the relatively complete construction "buys it" rather than "buys the". This effect would qualify as only an apparent support for the structures with a VP. The isolation effect, however, seems to be partially counteracted by another factor operating upon the PWC (man,koopt) vs. (koopt,boek). This factor could be the afore-mentioned surface distance effect.

Response bias. Still another interpretation in terms of *response bias* suggests itself after careful examination of the results of another PWC, viz. (de,man) vs. (het,boek). These results were given on Page 139. Under all twelve models, the choice probabilities for this PWC are .5 in accordance with our strongest syntactic intuitions. But, as Fodor et.al. (op.cit, p.366) argue: "It is not strange that, faced with the necessity of a choice, a subject casts about for a basis of choice and actually finds one". The observed choice frequencies, indeed, are 21 to 7, yielding a significant L-contribution of 7.33 to the various L-values. Being unwilling to make syntactic adjustments for this counterevidence, we think that it is interpretable as the result of a left-to-right response tendency. Although the order of the word pairs in the presentation of the PWCs was randomized anew for each subject in order to cope with response bias, this control has presumably failed. The PWCs in this study are formally dyads of stimulus pairs. They differ, however, from studies in which stimulus pairs comprise weights, lights or pitches in that the word pairs, though randomly ordered in the *presentations* of the PWCs, nevertheless possess a *natural order* in the sentence in which they occur. That this ordering seems to override the randomization effect is strongly suggested by the 2 x 2 frequency Table 6.8, in which the particular choices made -in terms of the left or right hand word pair- are paired off with order of presentation.

Table 6.8 Choice frequencies for the left or right hand word pair in relation to both orders of presentation

		word pair chosen:		
		left	right	total
order of presentation	(D,M) - (H,B)	9	2	11
	(H,B) - (D,M)	5	12	17
		14	14	28

The same "underlying" response bias may have been responsible for the asymmetric frequencies obtained for the PWC (man,koopt) vs. (koopt,boek), viz. 21 to 7, which are of the same order of magnitude as those for the above-mentioned PWC.

To summarize the above considerations, which of course are post hoc interpretations rather than explanations, we believe that symmetry of structure can be maintained as the most probable syntactic hypothesis. The violations can be interpreted by resort to non-syntactic factors. The observations p(KH.DK) > p(DK.KH) and p(KH.XY) > p(DK.XY) are supposed to reflect the isolation effect, whereas the observation p(MK.KB) > ½ is hypothesized to result from either superficial word distance or "underlying" response bias. These latter effects are likely to be at work only where syntactically indeterminate PWCs have to be judged in forced choice fashion and therefore do not necessarily extend to the comparisons of p(MK.XY) with p(KB.XY). Other possibilities such as an underlying CGV counteracted by response bias or word distance, or an underlying CGN counteracted by the isolation effect are much less plausible. Asymmetric structures would require *both* (man,koopt) *and* (de,koopt) to be more cohesive than *both* (koopt,boek) *and* (koopt,het) (CGN) or vice versa (CGV). Hence an extra requirement for CGN would be p(DK.KB) > ½ whereas the observed frequencies are 1 to 27. Likewise, the CGV-prediction p(MK.KH) < ½ violated by the observed frequencies 20 to 8, in spite of a possible isolation effect affecting (koopt,het). The symmetric structures, on the contrary, specify both (man,koopt) and (koopt,boek) as more cohesive than (de,koopt) and (koopt,het).

Besides the "variant" word pairs just discussed, the obstinate "invariant" word pair (de,boek) still stands in need of interpretation. In the previous section it was reported that this pair was considerably overpredicted by the three least unsuccessful models. For the PWC (de,boek) vs. (man,het), for instance, the observed frequencies are 5 to 23 "in favour of" (man,het). Under the restriction /de/ = /het/ and /man/ = /boek/, however, the weighted incompletednesses of these word pairs equal $a + n + v$ and the predictions are therefore p(DB.MH) = ½. Under our interpretation this is another instance of the isolation effect. In both pairs a noun is combined with an article from outside the dominating NP. In both pairs, moreover, the article and the noun are syntactically incompatible. The pairs differ, however, in the position of the article with respect to the noun. Since the article precedes the noun in the left hand pair it is likely to evoke an NP-interpretation, a possibility which, in our opinion, does not apply to the second pair. This evoked NP is, however, lexically ill-formed and might thus cause the subject to underrate the (D,B)-cohesion with respect to the (M,H)-cohesion, since *de boek* is incorrect Dutch. The same interpretation is presumably applicable to the other instances of the overprediction of (de,boek).

Choice theoretical aspects. Whereas the considerations concerning syntactic aspects of the cohesion models imply, at best, a few tentative pointers in favour of symmetry of structure, the indeterminacy is even greater with respect to the choice theoretical aspects. All three choice theories are represented among the three least unsuccessful models. Taken at face value, the results most favour the Thurstone approach. The CGR-Thurstone model yields the lowest goodness of fit measures (see Figure 6.2). The differences with CGP-Luce and CGP-Restle, however, are not substantial, even though their deterministic components render these models much more vulnerable. As a matter of fact, it is in this latter respect that the Luce and Restle approaches yield some problematic counterevidence.

Restle. In consideration of the division of PWCs into the four types given in Section 5.2 deterministic choice behaviour for type 3 PWCs (both word pairs incomplete, one incompleteness strictly included in the other) is a crucial feature of the Restle approach. Likewise, deterministic choice behaviour for the type 2 PWCs (one word pair complete, one incomplete) is crucial for the Luce approach (and, a fortiori, for the Restle approach). The results for the type 3 PWCs (according to CGP-Restle), with the more complete word pair mentioned first, are as follows: (man,koopt) vs. (man, het): 28 - 0; (man,koopt) vs. (man,boek): 26 - 2; (koopt,boek) vs. (de, boek): 28 - 0; (koopt,boek) vs. (man,boek): 25 - 3; (koopt,het) vs. (man, het): 28 - 0; (koopt,het) vs. (de,het): 26 - 2; (de,koopt) vs. (de,boek): 26 - 2; (de,koopt) vs. (de,het): 17 - 11. The last PWC especially exhibits a substantial violation of the Restle approach. In defense of the Restle approach, attention could be drawn to the correspondence of word categories occurring in the right hand member of the problematic PWC; the satisfactory choice frequencies obtained for the other type 3 PWCs should also be noted. Correspondence of word categories, however, also occurs in three other type 3 PWCs. Moreover, it can be argued that among the successful type 3 PWCs, there are four instances where the presumed (K,H)- and (D,B)-isolation effects may be operative in the direction of the deterministic predictions. Although we are inclined to eliminate the Restle approach from further consideration, a final decision will be suspended until the analysis of Experiment 2 with better controls for the isolation effect. Our prediction is that the introduction of controls for the isolation effect will undoubtedly discredit the Restle approach.

Luce, Thurstone. With respect to the type 2 PWCs, critical for CGP-Luce, it has already been observed (see the discussion of question 1) that the results for (de,man) vs. (man,koopt): 22 - 6; and (man,koopt) vs. (het,boek): 4 - 24 comprise non-negligible violations. On the basis of this counterevidence it is tempting to reject CGP-Luce in favour of the CGR-Thurstone model, which is very successful for the 18 type 2 PWCs (viz. those corresponding to cells in rows and columns 1 and 10 in Table 6.2, excepting cell(1,10)). The contributions from these PWCs to the overall L-values for CGR-Thurstone add up to 18.31, their MAD is .03. Nevertheless it can be seen that the problematic PWCs, together with certain other type 2 PWCs which, albeit to a lesser degree, fail to elicit uniform choice behaviour, almost all involve a word pair with two "content words" as the incomplete word pair. This suggests that a semantic factor such as the "amount of meaning contributed to the sentence" may be codetermining cohesion and may eventually modify the choice behaviour that is to be expected on syntactic grounds. In this connection one might wonder what else beside semantic factors could cause a pair like (man,boek) to be overrated in PWCs such as (man,boek) vs. (de,koopt) and (man,boek) vs. (man,het). In fact, these PWCs contribute 12.612 and 12.777 respectively to the overall L for CGR-Thurstone (resulting from underpredictions of -.322 and -.204), and 8.136 and 13.752 respectively to the overall L for CGP-Luce (underpredictions: -.253 and -.274). One might argue that the counterevidence reported for the type 2 PWCs is of the same semantic nature. This subtle dilemma between either the CGR-Thurstone model, which is preferable for reasons of parsimony, or an "integrated CGP-Luce model" with semantic codeterminants, cannot be resolved with the data of Experiment 1.

Suggestions for improving experimentation.

It will not be surprising that the analysis reported in the above pages and originally intended as an intermediate check of the data analytic aspects of this study, has been discontinued. The presence of so many confounding factors render the decisions concerning adequacy of the incompleteness principle and subtle choices between the probabilistic approaches and syntactic structures, impossibly difficult. Nevertheless, the experiment has taught us some valuable lessons. We shall therefore conclude this section on Experiment 1 by considering certain measures that might improve further experimentation in such a way that the confounding factors will be avoided where possible, or manipulated (and possibly explicitly accounted for in the model) where not.

The isolation effect might be avoided in both a specific and a more general way. Specifically, a careful examination could be carried out of all pairs of words to check whether their "togetherness" in a presentation could suggest an entity or relationship differing from that in the sentence. In such a case one would have to modify the sentence. As a matter of fact, Experiment 2 will use a modified version of the present sentence type, viz. *de man koopt een tas* (the man buys a bag). By replacing *het* with *een*, the most troublesome aspect in the design of Experiment 1, viz. the (koopt,het)-isolation effect, is expected to be eliminated. Moreover, the replacement of *boek* with *tas* ought to eliminate the (de,boek)-isolation effect. A more general way to escape the isolation effect is by completely avoiding separate presentations of the PWCs. This could be achieved, for instance, by indicating which PWC is to be judged by some graphical means. For an impression of how this can be accomplished, and in fact is realized in Experiment 2, see Figure 6.3 where the PWC (man,koopt) vs. (koopt,een) is indicated by means of arrows.

de man koopt een tas

Figure 6.3 Graphical indication of the stimulus PWC

Response bias seems to be limited to those PWCs where it is hard for the subject to reach a decision, as, for instance, we saw in the result for (de,man) vs. (het,boek). Since the left-to-right order *in the sentence* seems to override the randomly determined order *in the presentation*, we can see no other means for the avoidance of this effect, than by allowing the subject to give an equality judgment. The opportunity for giving equality judgments will also reduce a possible factor such as superficial *word distance*, which also seems to operate where subjects are rather indeterminate in their decisions.

6.3 EXPERIMENT 2

Problem. Experiment 2 was designed to shed a brighter light on the same set of concrete questions to which Experiment 1 was devoted. The results of Experiment 1 permitted a few tentative conclusions tending to support symmetry of syntactic structure (CGR and CGP) and seemed to favour the

Thurstone and Luce approaches over that of Restle. These tentative answers, however, had to be defended against a substantial amount of counterevidence, by interpreting the violations in terms of putative extraneous factors like word distance, isolation effects and response bias. In Experiment 2 several controls for these presumed effects are introduced. These have already been mentioned in the discussion of Experiment 1: the "arrow-method" of presentation of stimuli (Figure 6.3) for reducing the isolation effect; the opportunity for giving equality judgments in order to reduce factors likely to occur under forced choice: word distance and response bias; and another lexical version of the syntactic construction as an extra reduction of the isolation effect. The questions we now ask are the same as those enumerated in the problem section of Experiment 1. Our interest is in whether the tentative conclusions of Experiment 1 will find support under the new controls, whether these controls will allow for greater refinement of the conclusions, e.g. for a choice between CGR and CGP, as well as among the probabilistic approaches, and whether the controls will improve the overall goodness of fit measures to acceptable levels, or at least bring about a convincing improvement.

Method

Material, design. Computer generated stimulus material was used for the presentation of all 45 PWCs constructible from the 10 word pairs of the sentence *de man koopt een tas* (the man buys a bag). These PWCs, randomly ordered per subject and preceded by five training trials, were collected in small booklets of 50 pages, one PWC per page. Each PWC was indicated by means of the "arrow method" introduced in the discussion section of Experiment 1 (Figure 6.3). Which of the word pairs was to appear above and which below was decided on a random basis per subject.

Subjects, procedure. Forty nine subjects, first year students in Psychology at the University of Nijmegen, participated in the experiment. Their unpayed participation was in partial fulfilment of the first year curriculum requirements. The subjects were tested groupwise in an experimental session of about 15 minutes. The instruction was similar to that for Experiment 1 with the exception of an adjustment for the modified stimulus and response aspects. After reading the instruction the Ss had to work through the pages of the booklet, indicating on each page the pair that was felt to be most cohesive, or giving an equality judgment.

Results. After counting, the subjects' responses yielded the choice frequencies of Table 6.9. The cells give the numbers of subjects judging the corresponding row word pair as more cohesive than the corresponding column pair. Because of the equivalences, frequencies in symmetrically located cells do not necessarily add up to the total N of cases, the complement with respect to 49 being the number of equivalences.

Table 6.9 Choice frequencies for Experiment 2

	DM	DK	DE	DT	MK	ME	MT	KE	KT	ET
DM	-	47	47	45	36	46	46	45	38	8
DK	2	-	23	29	2	18	9	4	0	3
DE	0	7	-	12	1	7	1	4	2	1
DT	2	7	8	-	4	5	4	6	4	2
MK	10	47	46	45	-	45	41	39	14	6
ME	2	5	15	12	3	-	0	3	3	2
MT	2	28	31	39	4	41	-	21	6	3
KE	4	26	40	38	4	42	19	-	3	2
KT	8	46	43	44	9	44	35	42	-	5
ET	2	45	47	47	33	46	46	47	40	-

Equality judgments are alien to the choice theories in question, the predictions of which are exclusively in terms of inequalities. But an opportunity for giving equality judgments seems to be an indispensable methodological measure, as the current alternative, forced choice encounters the difficulties mentioned in the previous section. As an intermediate solution to this contradictory state of affairs it might seem a good idea to use a random number generator for deciding on the indeterminate PWCs and in this way to generate the "missing" portion of the data matrix. Ideally, however, one would have to repeat this procedure several times in order to prevent the analysis from becoming too dependent on the idiosyncracies of one such randomly completed data matrix. In order to avoid the enormous increase of the number of analyses that would result from this procedure, it was decided to divide the equalities evenly over the responses "more cohesive" and "less cohesive". This procedure yields the "empirical proportions" in

Table 6.10. As a preliminary step in the analysis of these data the L-test was applied to test the differences between the observed frequencies of Experiments 1 and 2 (the latter, of course, obtained by multiplying the en-

Table 6.10 Empirical choice proportions of Experiment 2, after an even division of the equalities over the other categories (decimal point omitted).

	DM	DK	DE	DT	MK	ME	MT	KE	KT	ET
DM	-	96	98	94	77	95	95	92	81	56
DK	04	-	66	72	04	63	31	28	03	07
DE	02	34	-	54	04	42	30	13	08	03
DT	06	28	46	-	08	43	14	17	09	04
MK	23	96	96	92	-	93	88	86	55	22
ME	05	37	58	57	07	-	08	10	08	05
MT	05	69	70	86	12	92	-	52	20	06
KE	08	72	87	83	14	90	48	-	10	04
KT	19	97	92	91	45	92	80	90	-	14
ET	44	93	97	96	78	95	94	96	86	-

tries in Table 6.10 by 49). A highly significant overall L-value of 101.156 ($df = 45$; $p < .0001$) was obtained, an indication that the modifications of the data gathering procedure had not been without consequences. The substitution of the word *tas* for *boek* in combination with the "arrow method" of stimulus presentation was introduced to avoid the (de,boek)-isolation effect. Actually, the nine PWCs with (de,boek) in Experiment 1 and with (de,tas) in Experiment 2 disproportionately contribute 41.508 to the overall L. Likewise, the 9 PWCs with (koopt,het) in Experiment 1 and with (koopt,een) in Experiment 2 contribute 25.407 to the overall L. This strongly suggests that the replacement of *het* by *een* in combination with the arrow method of presentation was effective in suppressing the (koopt,het)-isolation effect. Moreover, the freedom to given equality judgments seemed effective in reducing such factors as response bias. The choice frequencies obtained for the PWC (de, man) vs. (een,tas) no longer significantly contradict a choice probability of .5. As a consequence, after subtraction of the contributions from the PWC (de,man) vs. $(\{{\text{het,boek} \atop \text{een,tas}}\})$ and those in which $(\text{de},\{{\text{boek} \atop \text{tas}}\})$ and $(\text{koopt},\{{\text{het} \atop \text{een}}\})$

are involved, the residual L amounts to 39.711 with an associated residual df of 45 - 9 - 8 - 1 = 27; a value no longer significant at the 5% level.

Question 1 - deterministic versus probabilistic pairwise comparisons. In Experiment 2, the straightforward application of *Luce's choice theory* to the eight incomplete word pairs yields the unrelatedness values that are given in Table 6.11. Again, the values have been normalized in such a way that their raw sum of squares is unity. The resulting L-value is 23.21 (df = 21; p = .332); MAD (for probabilistic word pairs) = .033. So as far as the probabilistic PWCs are concerned, the linguistically unrestricted application of Luce's choice theory is again promising. Comparison with the scale values in Experiment 1 (Table 6.3) clearly shows that the relative position of the word pair (de,tas) in relation to the other scale values, considerably differs from that of (de,boek) in Experiment 1.

Table 6.11 Unrelatedness values resulting from the linguistically unrestricted application of Luce's choice theory to the eight "incomplete" word pairs

	de	man	koopt	een	tas
de	-	(.000)	.283	.547	.586
man		-	.021	.503	.104
koopt			-	.110	.025
een				-	(.000)
tas					-

The results obtained for the Luce-deterministic PWCs are similar to those obtained in Experiment 1. Twelve of the sixteen type 2 PWCs (see the entries in the 1st and 10th row and column of Table 6.9, cell (1,10) excepted) show choice frequencies acceptable within those limits of tolerance currently characterizing tests of deterministic models. Again, however, the PWCs (de, man) vs. (man,koopt): 36 - 9 and (man,koopt) vs. (een,tas): 6 - 33 represent notable exceptions, and so do the PWCs (de,man) vs. (koopt,tas) 38 - 8 and (koopt,tas) vs. (een,tas) 5 - 40. This is reflected by the MAD for the deterministic PWCs, which is .087. Our conclusion is that the present formulation of the Luce approach is not tenable, but whether this should imply outright rejection or an adjustment of the Luce models, will largely depend on the diagnostic analysis of their remaining aspects.

The Thurstone approach. The "unrelatedness values" obtained with the linguistically unrestricted application of the Thurstone model are given in Table 6.12. The overall L-value is 61.42 (df = 37; p = .007); MAD = .042. The results therefore differ from those obtained with the straightforward application of the Thurstone model in Experiment 1 and from those obtained with the probabilistic part of the Luce model in Experiment 2, in that the discrepancies between model and data are highly significant.

Table 6.12 Unrelatedness values obtained with the linguistically unrestricted application of Thurstone's choice theory.

	man	koopt	een	tas
de	0.000	1.924	2.304	2.300
man	-	0.575	2.228	1.412
koopt		-	1.435	0.702
een			-	0.022
tas				-

Since this unrestricted solution is evidently the limiting optimum obtainable through application under restrictions, the implication is that any linguistically restricted application of the Thurstone model is predestined to significantly differ from the data. No compensatory effect can be expected from the increase in the number of degrees of freedom in the restricted case, as the L reported is even significant when the number of degrees of freedom is 42, the number associated with the linguistic application. In spite of this finding we shall not exclusively rely on it for abandoning the Thurstone approach. Further investigation of the Luce approach shows a considerable worsening of the goodness of fit as soon as additional restrictions are introduced, and this soon leads to a consideration of the role of non-syntactic factors. For a fair comparison similar considerations in respect of the Thurstone approach should not be withheld.

The Restle approach. For the same reasons as those mentioned under Experiment 1 (Page 136), the evaluation *of the partitioning in deterministic and probabilistic* PWCs in the Restle model will be postponed until the investigation of the linguistically restricted applications of that approach.

Questions 2a, 2b and 2c - **probabilistic models under different restrictions**
The applications of the probabilistic models to the four syntactic options CGV, CGN, CGR and CGP with no other restrictions than positivity of weights (see question 2a) yielded the results given in the Tables A8 to A14 inclusive of the Appendix and represented graphically in Figures 6.4 and 6.5. Global inspection of the tables suffices to show that, as with Experiment 1, again none of the twelve models is particularly successful. All associated p-values are hypersignificant, their maximum being $.102 \times 10^{-4}$, viz. the p-value associated with CGP-Luce. Nevertheless there are clear improvements in the various goodness of fit measures despite the increase in the N of cases. For the least unsuccessful triple of models the relevant figures are as follows: CGP-Luce: L = 67.14 (df = 25) against 149.35 in Experiment 1, MAD = .072 against .099 in Experiment 1; CGP-Thurstone: L = 156.68 (df = 42) against 178.02 in Experiment 1, MAD = .077 against .094 in Experiment 1; CGR-Thurstone: L = 155.36 (df = 42) against 203.92 in Experiment 1, MAD = .075 against .105 in Experiment 1. A notable difference with the results of Experiment 1 is the relative drop of the CGR-Restle model. Whereas the L-value improved from 140.97 (df = 17) in Experiment 1 to 85.64 (df = 17), its overall MAD did not change much, viz. from .105 (Experiment 1) to .103 (Experiment 2). It must therefore be the case that here the probabilistic PWCs compensate for a deterioration of the deterministic component. Indeed, the MAD for deterministic PWCs only increases from .066 (Experiment 1) to .114 (Experiment 2) in accordance with what was expected to result from introducing controls for the isolation effect (see the discussion of Experiment 1, Page 152). The eight type 3 PWCs which according to CGP-Restle predict deterministic choice behaviour, yield the following deviations from unity: (de,koopt) vs. (de,een) .34, (de,koopt) vs. (de,tas) .28, (de,een) vs. (koopt,een) .13, (de,tas) vs. (koopt,tas) .09, (man,koopt) vs. (man,een) .07, (man,koopt) vs. (man,tas) .12, (man,een) vs. (koopt,een) .10 and (man,tas) vs. (koopt,tas) .20. Although for the sake of completeness the results for the Restle approach under the additional restrictions have been entered in the relevant tables and graphs, the observations above force us to exclude the Restle approach from further consideration: type 3 PWCs do not evoke deterministic choice behaviour.

As far as the Luce-approach is concerned it has already been established under "Question 1", that there are notable violations of the deterministic predictions with respect to the type 2 PWCs (PWCs with exactly one complete word pair). The MAD for the sixteen deterministic comparisons was no better than .087. The overall MAD-value (.072), however, is better than the MAD-values for CGP-Thurstone and CGR-Thurstone. This implies that the PWCs which are probabilistic according to CGP-Luce must have a lower MAD (viz. .064) than they have under CGP-Thurstone (.098) and under CGR-Thurstone (.103). In this connection, it is interesting to apply the well known z-transformation $z = \sqrt{2 \times \chi^2} - \sqrt{2 \times df - 1}$ to the L-values obtained in order to make them comparable in spite of differing numbers of degrees of freedom. The z-values

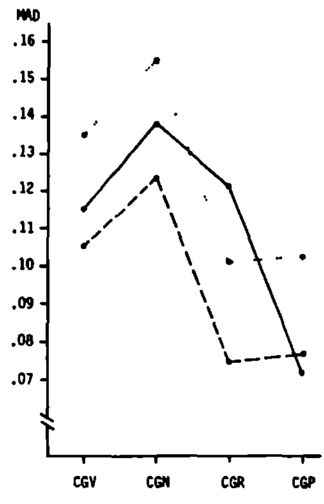


Figure 6.4 Experiment 2: overall MAD-values. No assumptions beyond positivity of weights (restriction a)

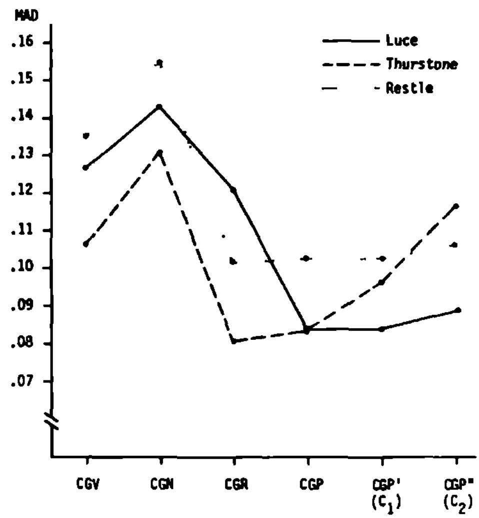


Figure 6.5 Experiment 2: overall MAD-values under restriction b: /de/ = /een/, /man/ = /tas/; for CGP moreover: restriction c1: /koopt/ ≥ /de/,/een/ (CGP') and restriction c2: /koopt/ ≥ /man/,/tas/ (CGP").

are: CGP-Luce: z = 4.59 (against 10.28 in Experiment 1); CGP-Thurstone: z = 8.59 (9.76 in Experiment 1); CGR-Thurstone: z = 8.52 (11.08 in Experiment 1). The significance of these differences, however, should not be overrated. Experiment 2 conforms to Experiment 1 in that, again, linguistically implausible weights are obtained. For the least unsuccesful triple of models, these weights are given in Table 6.13.

Table 6.13 Weights obtained for the three least unsuccessful models under restriction a (positivity of weights).

	/de/	/man/	/koopt/	/een/	/tas/
CGP-Luce	.071	1.000	.348	.082	.487
CGP-Thurstone	.672	1.454	.000	.619	1.148
CGR-Thurstone	.000	.704	1.255	.176	.694

Gross asymmetries of weights occur in CGP-Luce, cf. /man/ and /tas/; in CGP-Thurstone the /koopt/-weight reaches the imposed lower boundary of zero. Hence there is good reason for questioning the relative merits of these models under the introduction of the extra restrictions of questions 2b and 2c. Needless to say, the study of their behaviour under the extra restrictions again serves diagnostic rather than testing purposes.

The effects of the restrictions b and c on the twelve models are tabulated in Tables A11 to A14 inclusive in the Appendix and the overall MAD-values are depicted in Figure 6.5. In nine of the twelve representations under restriction b (/de/ = /een/; /man/ = /tas/), the three CGP-models excepted, the verb outweighs the nouns which, in turn, outweigh the articles. In other words, it was not necessary to test these models under either of the two versions of restriction c: c1 /verb/ $\geqslant$ /art/; c2 /verb/ $\geqslant$ /noun/. For the CGP-models, however, tests under these restrictions did have to be made. The MAD-values corresponding to these tests have been plotted above the points marked CGP' and CGP". Restricting attention to the models least unsuccessful under restriction a, it can be observed that CGR-Thurstone has the lowest overall MAD (.081) under restriction b, The corresponding L is 169.09 (z = 9.06). The weights are: /art/ = .048, /noun/ = .692 and /verb/ = 1.249, which clearly exhibits that this representation already meets restrictions c1 and c2. For CGP-Luce the relevant figures are: MAD = .084, L = 93.80 (z = 6.41) with weights /art/ = .109, /noun/ = 1.000* and /verb/ = .454. This solution already satisfies restriction c1. Under restriction c2 the following figures

* fixed in advance.

are obtained: MAD = .089, L = 110.01 (z = 7.55); weights: /art/ = .125, /noun/ = 1.000* and /verb/ = 1.000**. The CGP-Thurstone solution under restriction b is: /art/ = .642, /noun/ = 1.292 and /verb/ = .000**, with an associated MAD of .084, and L = 175.37 (z = 9.40). The results under c1 are: /art/ = .362, /noun/ = 1.048 and /verb/ = .362, MAD = .097, L = 212.98 (z = 11.31). Under restriction c2, finally, /art/ = .142, /noun/ = .733, /verb/ = .733, MAD = .117, L = 239.91 (z = 14.75).

In summary, we may conclude that the introduction of the restrictions b, c1 and c2 reduce the "leading triple" of models to a leading pair, viz. CGR-Thurstone and CGP-Luce. The overall MAD-values are in favour of the CGR-Thurstone model. A separate consideration of the Luce-probabilistic PWCs, however, indicates that, probabilistically, CGP-Luce is less unsuccessful than CGR-Thurstone. The high L-values, however, disqualify both models as unsatisfactory. Again, there is a choice to be made between integral rejection or adjustment of the models involved. A final decision will largely depend on the diagnosis to be given in the next section.

Question 3 - Sources of deficiencies

We shall now examine the counterevidence against the two competing models by considering in detail which PWCs and word pairs contribute most to the overall goodness of fit measures. A complete survey of how the various PWCs contribute to the overall L- and MAD-values for CGP-Luce (under restriction c2) and CGR-Thurstone (under restriction b) is given in the Appendix, Tables A16 and A17 respectively. As far as the word pairs are concerned, three interpretative indices are tabulated in Tables A18 and A19 of the Appendix, again, for CGP-Luce and CGR-Thurstone respectively. For each word pair (X,Y) these indices are: (a) its contribution to the overall L-value, defined as the sum of the contributions of the PWCs in which the word pair (X,Y) is involved as one of the members; (b) the MAD for the PWCs in which the word pair (X,Y) is involved; (c) its overprediction, O(X,Y), the definition of which has been given on Page 147 as the mean of the algebraic deviations $\hat{p}(XY,\ldots) - p(XY,\ldots)$.

In support of the interpretability of these figures, the data analytic procedure was supplied with an iterative interpretative subroutine which

* fixed in advance

** imposed lower boundary

works as follows. In each of a series of iterations it is determined which PWC contributes most to the overall L-values. It is recorded whether this contribution is due to over or underprediction of the left hand pair with respect to the right hand pair. This PWC is then eliminated from the analysis. The residual L is calculated and "tested" for significance (with the number of degrees of freedom, of course, appropriately adjusted). The subroutine continues with a new iteration unless the L is no longer significant at a certain predetermined level (say, of 10%). The interpretative subroutine is then completed with a similar successive elimination procedure for the *word pairs*.

For CGP-Luce, this part of the analysis yielded the results given in Tables 6.14 and 6.15.

Table 6.14 Successive elimination of PWCs for CGP-Luce.

iteration	maximally contributing PWC	contribution to L	overprediction	residual L	residual df	p
1	ME - MT	21.865	.289	88.125	26	.000
2	ME - KE	13.037	.218	75.087	25	.000
3	DT - MT	12.568	.227	62.519	24	.000
4	DK - MT	12.407	.249	50.113	23	.001
5	DK - KE	10.240	.224	39.873	22	.011
6	MT - KT	5.804	-.113	34.069	21	.035
7	DT - KE	5.397	.146	28.672	20	.094
8	DK - KT	4.359	.080	24.312	19	.184

Table 6.15 Successive elimination of word pairs for CGP-Luce

iteration	maximally contributing word pair	contribution to L	overprediction	residual L	residual df	p
1	MT	56.888	-.133	53.102	20	.000
2	KE	33.299	-.106	19.802	14	.136

The analysis clearly localizes the violations of the probabilistic* component of the CGP-Luce model in those PWCs in which (man,tas) or (koopt,een) are

* Violations of the deterministic component have already been discussed at the beginning of the results section.

involved as one of the pairs. Moreover, the analysis indicates that the empirical relatednesses of these pairs are underpredicted by the model. In the case of the pair (man,tas) we presume that a semantic factor, say, amount of contribution to the meaning of the sentence raises its relatedness above what is to be expected on the basis of syntactic incompleteness only. This point will be reconsidered in the discussion section. With respect to the pair (koopt,een), it must be recognized that here the violations concern the very syntactic aspects of the interpretation theory. On the one hand, Experiment 2 corresponds to Experiment 1 in globally supporting "symmetry of syntactic structure" over asymmetry. On the other hand, the underestimation of the pair (koopt,een), which does not affect the pair (de,koopt), reflects an empirical asymmetry in the data which is inconsistent with one of the consequences of symmetry of structure (see Page 145). We shall return to this point in the discussion section as well.

For the CGR-Thurstone model the successive elimination procedure is reported in Tables 6.16 and 6.17.

Table 6.16 Successive elimination of PWCs for CGR-Thurstone

iteration	maximally contributing PWC	contribution to L	overprediction	residual L	residual df	p
1	DK - KT	19.186	.229	149.900	43	.000
2	DE - DT	17.402	-.281	132.498	42	.000
3	DK - MK	16.935	.219	115.564	41	.000
4	DK - DE	16.628	.221	98.936	40	.000
5	DK - KE	10.239	.224	88.697	39	.000
6	ME - KE	10.147	.187	78.550	38	.000
7	ME - MT	10.130	.178	68.419	37	.001
8	MK - MT	7.807	-.167	60.612	36	.006
9	KE - KT	7.640	.158	52.972	35	.026
10	DE - ME	5.810	-.159	47.162	34	.066
11	DK - MT	5.127	.159	42.035	33	.135

Since the CGR-Thurstone model shares the equality of (de,koopt) and (koopt, een) with the CGP-Luce model, it conforms to the expectation that the empirical asymmetry affecting these pairs is reflected by the above analysis as well.

The results, however, differ from those obtained for CGP-Luce in suggesting that the violations reflect overprediction of the cohesion of (de,koopt) rather than underprediction of (koopt,een). A second difference with the

Table 6.17 Successive elimination of word pairs for CGR-Thurstone

iteration	maximally contributing word pair	contribution to L	overprediction	residual L	residual df	p
1	DK	70.585	.128	98.501	35	.000
2	DE	34.909	-.082	63.559	27	.000
3	KE	26.336	.002	37.255	20	.011
4	MT	25.298	.004	11.957	14	.610

CGP-Luce analysis is the role of the pair (de,een). The syntactic predictions of CGR-Thurstone, on the average, underpredict its empirical relatedness. It is perhaps tempting to ascribe the counterevidence to the correspondence of word categories. This suggestion, however, is not supported by a comparable underprediction of the pair (man,tas), which plays a very modest role in the list of violations of CGR-Thurstone as opposed to its place in the CGP-Luce analysis. The same can be said of the pair (koopt,een) which does appear on the list of PWCs eliminated, but with alternating over and underpredictions which almost cancel each other. Generally, the pattern of violations of CGR-Thurstone is hard to interpret in terms of alternative (i.e. non-syntactic) factors. We shall return to this point in the discussion section.

Discussion and some further explorations

Syntactic aspects. On the basis of the goodness of fit measures two models distinguish themselves as "least unsuccessful": CGR-Thurstone and CGP-Luce. Experiment 2 confirms the main result of Experiment 1: for the sentence studied, the symmetric structures give a better account of cohesion judgments than the asymmetric ones. It is interesting to relate this result to the syntactic issues discussed in Chapter 3. CGP-Luce can be characterized as a probabilistic application of the incompleteness principle to the *set of substructures* implied by the *D-structure* of the sample sentence. CGR-Thurstone can be conceived of in a double sense. It can be seen as a probabilistic application of the incompleteness principle to the *set of complete substructures* implied by the same *D-structure*, or to a particular option according to the *C-formalism*, viz. a C-grammar whose first rewriting rule is S → NP V NP. This C-grammar, however, is of degree one (see Chapter 3) and accordingly it belongs to the trivial subclass of C-grammars whose members can be put into type 1 correspondence with D-structures. The C-formalism's extra capacity for specifying grammars of a degree which exceeds unity leads -for the sentence type used here- to the asymmetric models CGV or CGN, whose incorporation into the cohesion model turned out to be very unsuccessful. For the sentence in question, the results therefore advocate either the D-grammar (CGP) or a C-grammar (CGR) which, in lacking those features that typically distinguish C- from D-grammars, can be imitated by a corresponding D-grammar. The results, in other words, favour a preference for the D-grammar.

Empirical asymmetries. However, some of those consequences of symmetry of structure mentioned on Page 145 have again been violated, albeit to a lesser degree in comparison with Experiment 1. The violations concern the predicted equality of (de,koopt) and (koopt,een), implying (i) p(DK.KE) = ½ and (ii) p(DK.XY) = p(KE.XY) for all word pairs (X,Y) ≠ (D,K) or (K,E). The PWC (de, koopt) vs. (koopt,een), however, with choice frequencies 4 - 26 and adjusted proportions .27 - .73, contributes 10.240 to the overall L-values of both competing models. This empirical asymmetry recurs in the relevant cells of the rows (columns) headed DK and KE in Table 6.9. Obviously, the control measures introduced to reduce the (koopt,het)-isolation effect of Experiment 1, although partially effective (see Pages 157 and 158), did not fully reduce the asymmetry of the pairs (de,koopt) and (koopt,een). Since the incorporation of CGV as the syntactic component of the model globally worsens the goodness of fit, we still adhere either to CGR or CGP. This decision again confronts us with the problem of explaining how symmetric structures are able to yield asymmetric judgments for the PWCs under the scope of the afore-mentioned implications (i) and (ii). Stated more concretely, in which respects thus far neglected in the cohesion model, do the problematic pairs (de,koopt) and (koopt,een) differ? Three salient differences can be mentioned. Firstly, the pair (de,koopt) precedes the pair (koopt,een) in the sentence used. Secondly, the superficial word distance in the former pair exceeds that in the latter pair, whose words, unlike the other's are concatenated. Thirdly, in the first pair the article precedes the verb, in the second pair this order is reversed.

As far as the first point is concerned, *precedence* in the sentence is sometimes (see Experiment 1) reflected in a left-to-right response bias in the choice frequencies. An opposite effect, however, is observed for the PWC (de, koopt) vs. (koopt,een). The first point is therefore unable to serve as an explanation in this context. With respect to the second point, the considerations on Page 149 should be recalled, that thus far, there is not much evidence for the effectiveness of *superficial word distance*. Moreover, the admission of equality judgments was intended to reduce the effects of such a possible factor. Nevertheless, though we are doubtful about the influence of word distance, the results obtained oblige us to keep this factor in consideration. The third point, *difference of word order,* is related to another distinctive characteristic which we shall henceforth refer to as *concatenability*. The first pair, (de,koopt) does not appear as a subsequence in the sentence and can never be a subsequence of any sentence* . The transition probability associated with (de,koopt), so to speak, is zero. On the other hand, the transition probability of (koopt,een) is definitely greater than zero, since (koopt,een) is a possible substring in Dutch sentences (in fact, it is a substring in the sentence dealt with here). *Koopt* and *een* are, in our terminology, concatenable as opposed to *de* and *koopt*.

Suggestions for further experimentation. An interesting way to determine which, if any, of the afore-mentioned factors are able to account for the empirival inequality of (de,koopt) and (koopt,een) is to introduce transformed versions of the experimental sentence ("transformed" and "transformation" will be used in a loose, non-technical way, which does not involve commitment to a particular theory of syntax). It should be recognized that an explanation of the observed asymmetry in terms of word distance or concatenability implies that factors, closely related to word order and hitherto neglected, are co-determinants of cohesion. And since word order varies with transformations, cohesion judgments should also vary to the degree that transformations affect word distance and concatenability. As a byproduct of the study of transformed versions of the experimental sentence, we may obtain evidence about the generally assumed role of underlying relations in the determination of cohesion.

From this point of view, it is interesting to consider how the

* With the exception of this one itself.

choice behaviour for the problematic PWC (de,koopt) versus (koopt,een) might be affected if we change the experimental sentence into its interrogative and object clause versions. In Dutch, the interrogative and the object clause versions are "*koopt de man een tas*" (does the man buy a bag) and "hij zegt *dat de man een tas koopt*" (he says that the man is buying a bag) respectively. As far as *word distance* is concerned, we see that the object clause version corresponds to the simple declarative version in that *de* and *koopt* are further apart than *een* en *koopt*, whereas in the interrogative version it is the other way round. On the basis of word distance, in other words, for the object clause version we would expect: $p(DK.KE) < \frac{1}{2}$ and $p(DK.XY) < p(KE.XY)$ (with (X,Y) varying over all word pairs except DK and KE), and for the interrogative version $p(DK.KE) > \frac{1}{2}$ and $p(DK.XY) > p(KE.XY)$. The underprediction of (koopt,een), or overprediction of (de, koopt) should also hold for the object clause version. In the interrogative version, on the contrary, overprediction of (koopt,een), or underprediction of (de,koopt) ought to be expected.

As far as *concatenability* is concerned, both the interrogative and object clause versions deviate from the simple declarative version in which only one pair, (koopt,een), qualifies as a possible subsequence, in contrast with (de,koopt). In the interrogative version *both* pairs are possible sequences, in the object clause version neither is. Accordingly, for both transformed versions the predictions would be $p(DK.KE) = \frac{1}{2}$ and $p(DK.XY) = = p(KE.XY)$ (with $(X,Y) \neq (D,K)$ or (K,E)). So if concatenability were indeed the exclusive determinant of the asymmetry in the data for the simple declarative version, this asymmetry would be expected to disappear under an interrogative and object clause transformation. As attempt to study the asymmetry problem in terms of the variables word distance and concatenability will form one of the topics of Chapter 7. There we shall investigate the interrogative and object clause versions of the model sentence using the same experimental and data analytic methodology as used in Experiment 2.

Choice theoretical aspects. With respect to the second aspect of the interpretation theory, viz. the probabilistic choice theory to be incorporated, Experiment 2 has provided us with some additional refinements. First of all, it has definitively disqualified the Restle approach, since its demarcation between probabilistic and deterministic components is empirically untenable: type 3 PWCs ought, but undoubtedly fail to elicit deterministic choice behaviour. As far as the Luce and the Thurstone approaches are concerned, the results of Experiment 2 are less compelling; nevertheless the comparison is in favour of the Luce approach. The goodness of fit measures for the least unsuccessful restricted applications, CGR-Thurstone and CGP-Luce turned out to be not very dissimilar. The small disadvantage of CGP-Luce, as far as the overall MAD-values are concerned (see Figure 6.5), appeared to be mainly the consequence of strong violations of deterministic PWCs. Probabilistically, the Luce approach yields a better fitting solution than the Thurstone approach, both in the linguistically restricted applications (CGP-Luce vs. CGR-Thurstone) and in the unrestricted cases. As a strong argument against the Thurstone approach we have observed that even its unrestricted and thus most lenient application yielded a highly significant L-value. This did not seem to be the case for the authentic application of Luce's choice theory. Another, and from the "realistic" point of view very serious argument against the Thurstone approach, is that its optimal linguistic application, viz. CGR-Thurstone, yields a pattern of deficiencies that is hardly interpretable in terms of non-syntactic factors. On the basis of these considerations we are inclined to opt for the Luce approach. Nevertheless, some qualifications should accompany this conclusion.

Some qualifications concerning the preference for CGP-Luce. The first qualification concerns the generality of the conclusion. The moderate "superiority" of CGP-Luce over CGR-Thurstone is established for only one sentence type and should, therefore, be tested for other sentence types as well. The second qualification relates to some conditions that go with the conclusion. The adequacy of the CGP-Luce model is made conditional on various a posteriori interpretations of violations in terms of certain order-dependent and semantic factors. The next chapter will, among other things, devote itself to a study of the tenability of these interpretations. The third qualification relates to a further notable exception that should not be overlooked: the violations of the deterministic predictions in some of the type 2 PWCs. The present formulation of the Luce approach says that a comparison between complete and incomplete word pairs will result in a deterministic choice in favour of the complete pair, irrespective of the incompleteness value of the incomplete pair. The observed choice proportions, however, in the cells 2 to 9 inclusive of the first and tenth rows of Table 6.10 contradict these predictions and suggest a monotonic relation with the incompletenesses of the relevant column pairs. The implication is that if we decide on Luce's choice theory as the most appropriate probabilistic rule, adjustments should be made allowing for the deduction of probabilistic predictions for those PWCs which, according to the present formulation, are type 2. For the Luce approach (unlike the Restle approach, for reasons to be considered below), we see a relatively simple and general technique for such an adjustment.

Adjustment of CGP-Luce. This technique amounts to turning all type 2 PWCs into type 3 PWCs by introducing an extra "dummy" element in all word pairs' incompletenesses. This prevents the pairing of zero and non-zero incompletenesses and makes all PWCs probabilistic. The extra dummy element might be assigned the status of a "technical adjustment" (see Section 2.4.1) although a syntactic interpretation of the following kind might be adopted. One can think of an experimental situation in which PWCs are presented whose pairs are not only word pairs but constituent pairs as well. Intuitively, the perceived cohesiveness of the constituent pair (de man, koopt een tas) would be greater than that of all other complete pairs, because of its completeness in an additional sense: it is the only pair that is exhaustive with respect to the whole sentence. Actually, such a result was obtained in the experiment reported in Chapter 4 (see Table 4.2), where the constituent pair (de jongen, slaat een vriend) received the highest average rating. One way to account for this would be to add an extra column to the structure matrix of the sentence, whose entries are 1's for all word and constituent pairs lacking "sentencehood". The "technically adjusted" structure matrix suggested above might then be conceived of as a submatrix of this extended supermatrix. For the data of Experiment 2, application of the extra constant approach to the CGP-structure (under restriction c2) yields a promising step forward as will appear from the following comparison with the CGP-Luce solution discussed thus far.

Although the number of probabilistic PWCs increases from 29 to 45 the increase of the L is small: 119.87 ($df = 42$, $p < .0001$, $z = 6.37$) against 110.01 ($df = 27$, $p < .0001$, $z = 7.55$) in the original solution. Due to the elimination of deterministic PWCs, the MAD-value decreases remarkably from .089 to .067 (for comparison, the MAD associated with the CGR-Thurstone solution was .081). The parameters obtained with the extra constant approach are as follows: /art/ = .104, /noun/ = /verb/ = 1 (fixed), the extra constant, $\kappa = .055$. (For comparison, the parameters obtained with the original CGP-Luce solution were /art/ = .125 and /noun/ = /verb/ = 1 (fixed)).

For the Restle model we do not see a similar general and simple technique for the probabilization of the deterministic PWCs. Since the choice probabilities for a particular PWC depend upon the unique elements of the incompletenesses with respect to each other, in this model they remain unaffected by the introduction of a common extra element to the word pairs' incompletenesses. The extra element would be an element of the intersection of the incompletenesses involved in the PWC and not of their set differences. Probabilization of the type 2 and type 3 PWCs in terms of the Restle model would, therefore, require the introduction of unique elements in those incompletenesses which are included in others. We do not see any theoretical basis for such an ad hoc adjustment.

Non-syntactic aspects. A third point to be discussed relates to the non-syntactic determinants of the cohesion judgments and the means of accounting for them in the interpretation theory. In the diagnostic section it was found that the word pair (man,tas) was systematically underpredicted by the CGP-Luce model. Our interpretation ascribes this to a semantic factor. Since the pair (man,tas) contains two content words, it contributes relatively largely to the meaning of the sentence, a fact which, especially in the more indeterminate cases, might tend to reinforce the intuited relatedness. If our interpretation holds, this relatedness will be underpredicted by a mere syntactically based prediction. If an additional semantic factor is at work, this will not only affect the overall L-value, but also differentially affect the results for the other PWCs. The syntactic incompleteness of (man,tas) is {de,koopt,een}. Under the restriction that /de/ = /een/ = /art/, the weighted incompleteness is 2 x /art/ + /verb/. In our interpretation, the intuited unrelatedness of (man,tas) is smaller than is indicated by this weighted syntactic incompleteness. We shall express this by writing the "true" unrelatedness as u(M,T) = /I(M,T)/ - sem, where *sem* reflects the reduction due to the semantic factor. In handling a model, however, in which u(M,T) = /I(M,T)/, the accompanying data analysis will tend to reduce /I(M,T)/, i.e. either /art/ or /verb/ or both, possibly with undesired global consequences for other aspects of the model. It is therefore essential to somehow control the (man,tas) effect, even for the fundamental goals of this study.

Since we can see no satisfactory technique for controlling this semantic factor experimentally, e.g. by elimination or manipulation, as for the isolation effect or word distance, we prefer an explicit account of it in the interpretation theory. The way to accomplish this has already been suggested above: the introduction of an extra (sub-) additive parameter *sem* for the pair (man,tas) in the structure matrix. The effect intended is that the underprediction of the relatedness of (man,tas) will be reduced by subtracting a semantically interpreted constant from the syntactic incompleteness of this pair. For all PWCs involving (man,tas) this will result in an increase of the probabilities p(MT.XY).

Application of this proposal results in a noteworthy improvement in the cohesion model. The following parameter-values are obtained: /art/ = .111, /noun/ = 1 (fixed), /verb/ = 1.696 (!), sem = -1.307, κ (i.e. the additive constant for all word pairs' incompletenesses) = .053. The associated L-value is 75.931 (df = 41, p = .967 x 10^{-3}), MAD = .053. The diagnostic subroutine reveals that under this approach one source of violations remains, viz. the asymmetry of (de,koopt) and (koopt,een). As already mentioned, the task of the next chapter will be to reveal whether it is justifiable to explain this deficiency away in terms of word distance or concatenability. The most remarkable aspect of this analysis is that it is the first instance in this study in which an application of the CGP-Luce model "spontaneously" yields a verb parameter exceeding the noun parameter. Until now, this has been accomplished by imposing the restriction /verb/ $\geqslant$ /noun/ on the structure matrix in the way, described in Chapter 5, Section 5.3. Without the restriction (see Page 162) the weights for /noun/ and /verb/ were 1.000 and .454 respectively, but after imposition of the restriction the verb parameter took the value of the noun parameter, viz. unity, the smallest value possible under the restriction. From the point of view of the semantically adjusted analysis given above, it seems that in all these cases the /verb/ parameter assumed unrealistic values because such a reduction was the only way to absorb the semantic effect associated with the pair (man,tas) within the syntactic model.

Adjustment of the incompleteness principle. A final remark deals with the modification of the incompleteness principle, implied by the above analysis. This analysis has given a neat example of how the *realization problem* (cf. Chapter 1) takes the form of embedding a linguistic theory in a more integrated theory of linguistic performance. In this instance the embedding is accomplished in the following way, which will also be adopted in the next chapter. A distinction is made between the syntactic incompleteness /I(X,Y)/ of a word pair (X,Y) and its *unrelatedness* u(X,Y) of which /I(X,Y)/ is a component part. The syntactic incompletenesses continue to be specified by means of the incompleteness matrices introduced in Chapter 5, with rows, corresponding to word pairs and columns to the "syntactic" parameters. By introducing additional column vectors to this incompleteness matrix, incorporating the word pairs' specifications with respect to non-syntactic determinants, we obtain the unrelatedness "supermatrix". Luce's choice theory is now applied to the unrelatednesses in the obvious way: p(XY.WZ) = u(W,Z) : (u(W,Z) + u(X,Y)). The data analysis then proceeds in the same formal way as described in Section 5.2.

Summary. In summarizing the analyses of the data of Experiment 2, we may tentatively conclude that, as far as syntax is concerned, the evidence supports the incorporation of CGP, i.e. the D-grammar into the cohesion model. Apart from obvious limitations on external validity, i.e. generalizability of conclusions beyond subjects, linguistic material and experimental conditions selected, the conclusion must remain tentative for reasons of internal validity. The conclusion is conditional upon the adequacy of the other aspects of the interpretation theory: the incompleteness principle and the probabilistic interpretation axiom, relating choice probabilities to the ratio of weighted incompletenesses. A disappointing goodness of fit of the CGP-Luce model forced us to analyze the counterevidence against it in detail. On the basis of this analysis it was possible to reduce the violations to three sources of deficiencies: (i) non-deterministic choice behaviour for type 2 PWCs, (ii) underpredictions of the word pair (man,tas) and (iii) inequality of the relatednesses of (koopt,een) and (de,koopt). Two relatively simple adjustments of the cohesion model have been proposed for coping with the first two sources of deficiencies. The introduction of an additive constant to all weighted incompletenesses eliminated the first category of violations. The addition of a unique (sub)additive, semantically interpreted parameter to the syntactic incompleteness of (man,tas) eliminated the second category of violations. With respect to the third category of deficiencies, interpretations in terms of word distance and concatenability were considered. It was argued that an experimental study of "transformed" versions of the experimental sentence of Experiment 2 in which these presumed causes are varied, may be helpful in detecting the cause of the asymmetry. This experimental study will form the subject of Chapter 7.

CHAPTER 7 / WORD ORDER AND COHESION

7.1 *EXPERIMENT 3*

Introduction. Experiment 2 has reinforced our preference for the "symmetric structures", CGP and CGR, over the asymmetric ones, CGV and CGN. In this connection, however, a remaining puzzle was how underlying symmetric structures can yield asymmetric judgments for the word pairs (de,koopt) versus (koopt,een). Interpretations were given in terms of *word distance* and *concatenability*. It was argued that an experimental study, using the interrogative and object clause versions of the specimen sentence, might help us find out whether any of these factors are responsible for the asymmetry. An illustration (Figure 7.1) depicts what happens to the problematic pairs under the interrogative and object clause "transformations".

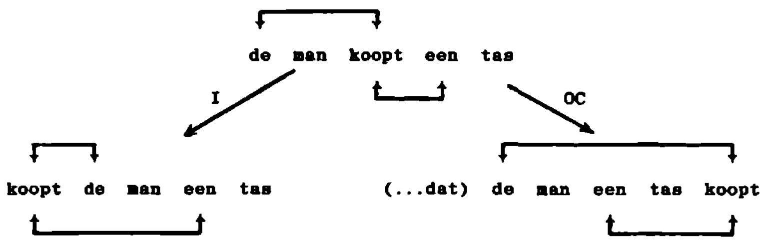


Figure 7.1 Changes in the PWC (de,koopt) vs. (koopt,een) under replacement of *de man koopt een tas* by its interrogative (I) and object clause (OC) versions

In the interrogative version the words *koopt* and *een* are deconcatenated but remain in the "declarative"order, in which they are concatenable. The order of the pair (de,koopt) has become (koopt,de). In this order the words are concatenable (in fact, they *are* concetenated), unlike the case for the simple declarative version. As far as superficial word distance is concerned, d(koopt,de) < d(koopt,een), in contrast with the declarative version where d(de,koopt) > d(koopt,een). Under the concatenability interpretation and the hypothesis of underlying syntactic symmetry, the original asymmetry (de, koopt) $\leq_c$ (koopt,een) ought therefore to disappear in the interrogative version. Under the distance interpretation, inversion of the asymmetry in the interrogative version would be expected. In the object clause version both (de,koopt) and (een,koopt) are non-concatenable. As far as distance is concerned d(de,koopt) > d(een,koopt). Hence were concatenability the exclusive cause of the original asymmetry, equality is to be expected for the object clause version. Under the distance interpretation, however, the original asymmetry ought to recur. The predictions following under the alternative interpretations for both transformed versions have been summarized in Table 7.1 in terms of the choice probabilities p(DK.KE).

The data of Experiment 2 do not reveal a corresponding significant asymmetry for the PWC (man,koopt) versus (koopt,tas). Nevertheless, there is a small inequality in the choice frequencies: after adjustment these choice frequencies become 27 and 22 respectively. In view of the above interpretations of the (de,koopt) vs. (koopt,een)-asymmetry the possibility remains that here the test was of insufficient power to detect a "genuine" asymmetry. An additional consideration is the evidence of much past experimentation (see, for instance, Experiment 1), showing a slight dominance of the subject-N to verb relatedness over the object-N to verb relatedness. If, on these grounds, significance is assigned to the (man,koopt) versus (koopt, tas) inequality in the declarative version's data, additional expectations for the transformed versions can be derived under the word distance and concatenability interpretations. Figure 7.2 illustrates what happens under the interrogative and object clause tranformations to these problematic pairs. In the interrogative version the order of the words *man* and *koopt*, previously concatenated in the declarative version, becomes reversed; this

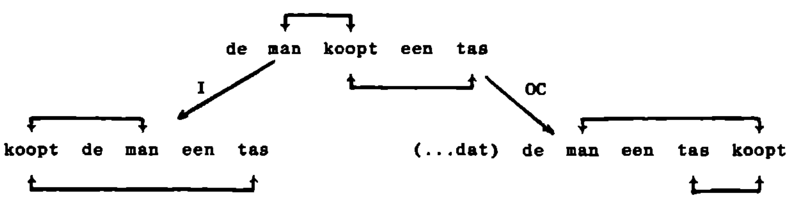


Figure 7.2 Changes in the PWC (man,koopt) versus (koopt,tas) under replacement of *de man koopt een tas* by its interrogative (I) and object clause (OC) versions

renders the words inconcatenable as is also the case with *koopt* and *tas*. The superficial word distance of *koopt* and *man* becomes increased but is still smaller than that between *koopt* and *tas*. So if there is a genuine (man,koopt) versus (koopt,tas) asymmetry, then it ought to recur in the interrogative version when distance is decisive. It will, however, turn into an equality under the concatenability interpretation.

In the object clause version's word order, the pair (man,koopt) remains concatenable, whereas the non-concatenable pair (koopt,tas) is permuted into the concatenable, and concatenated, (tas,koopt). The increased word distance of (man,koopt) now exceeds that of the pair (tas,koopt). According to the distance interpretation, a reversed asymmetry viz. (man,koopt) $<_s$ (tas, koopt) ought to be expected; under the concatenability interpretation, the asymmetry would have to change into an equality. Table 7.1 summarizes all these predictions in terms of the choice probabilities p(MK.KT) in addition to the afore-mentioned p(DK.KE)-values.

In the experimental study here described as Experiment 3, we shall examine whether the pattern of choice frequencies for the above transformed versions of the sample sentence is consistent with the expectations under either the distance or the concatenability interpretation. Moreover, since the above interpretations imply that superficial aspects of structure, co-

Figure 7.1 Summary of predictions p(DK.KE) and p(MK.KT) under the concatenability and word distance interpretations for the declarative, interrogative and object clause versions of the experimental sentence

		distance	concatenability
p(DK.KE)	declarative version	< .5	< .5
	interrogative version	> .5	= .5
	object clause version	< .5	= .5
p(MK.KT)	declarative version	> .5	> .5
	interrogative version	> .5	= .5
	object clause version	< .5	= .5

varying with transformations, also determine cohesion, it can not be excluded a priori that transformed versions might favour different syntactic structures. Therefore, the full collection of syntactic options dealt with thus far will be reconsidered -under the Luce approach- in connection with these transformed versions. This aspect of the data analysis checks the assumption of symmetry of syntactic structure underlying the predictions in this sub-section as summarized in Table 7.1.

Method

As far as the experimental methodology is concerned, Experiment 3 is an exact replication of Experiment 2. For details of design, organization of the material and procedure we therefore refer the reader to the description of Experiment 2. Two independent samples of subjects, 50 for the interrogative version and 49 for the object clause version, were taken from the same population as used in Experiment 2.

Results

The choice frequencies and the adjusted choice proportions obtained for Experiment 3, are presented in Tables 7.2 to 7.5 inclusive in the same way as those for Experiment 2 (see Tables 6.9 and 6.10). Tables 7.2 and 7.3 concern

Table 7.2 Choice frequencies for Experiment 3, interrogative version

	DM	DK	DE	DT	MK	ME	MT	KE	KT	ET
DM	-	49	48	49	45	47	46	49	39	5
DK	1	-	19	25	1	11	6	5	2	1
DE	2	9	-	8	2	10	1	3	1	0
DT	1	5	9	-	5	5	3	2	4	0
MK	5	43	43	43	-	44	31	37	14	7
ME	3	2	17	8	2	-	2	3	0	1
MT	4	32	36	42	4	40	-	31	5	5
KE	1	27	39	29	7	31	12	-	1	1
KT	8	46	46	44	22	46	37	47	-	9
ET	1	49	50	49	43	49	45	49	39	0

the interrogative version, Tables 7.4 and 7.5 the object clause version of the sample sentence. For reasons of comparability, the word pairs associated with the rows and columns of the matrices involved are presented in the same order as that for the declarative version. As a first step in the analysis of the data, the L-test was applied to examine the differences between the choice frequencies of Experiment 3 (i.e. the proportions of the Tables 7.3

Table 7.3 "Observed" choice proportions for Experiment 3, interrogative version; equalities evenly divided over categories "greater than" and "less than"; decimal points omitted

	DM	DK	DE	DT	MK	ME	MT	KE	KT	ET
DM	-	98	96	98	90	94	92	98	81	54
DK	02	-	60	70	08	59	24	28	06	02
DE	04	40	-	49	09	43	15	14	05	00
DT	02	30	51	-	12	47	11	23	10	01
MK	10	92	91	88	-	92	77	80	42	14
ME	06	41	57	53	08	-	12	22	04	02
MT	08	76	85	89	23	88	-	69	18	10
KE	02	72	86	77	20	78	31	-	04	02
KT	19	94	95	90	58	96	82	96	-	20
ET	46	98	100	99	86	98	90	98	80	-

Table 7.4 Choice frequencies for Experiment 3, object clause version

	DM	DK	DE	DT	MK	ME	MT	KE	KT	ET
DM	-	48	48	43	35	48	40	45	38	4
DK	1	-	24	23	0	18	8	15	13	1
DE	1	9	-	9	2	10	2	7	2	0
DT	2	6	13	-	8	8	8	7	8	3
MK	7	49	46	41	-	46	37	47	37	8
ME	0	5	14	7	2	-	4	6	2	0
MT	6	35	32	35	4	35	-	30	18	6
KE	2	11	26	23	1	26	14	-	4	5
KT	7	31	40	37	2	37	19	41	-	6
ET	2	47	49	45	32	48	41	41	42	-

and 7.5 appropriately multiplied by the number of subjects) and those for the declarative version, using the latter as the basis of prediction. For the interrogative version's data an L of 66.440 (df = 45, p = .020) was

obtained and a MAD value of .046. For the object clause version these figures are: L = 208.374 (p < .0001), MAD = .078.

Table 7.5 "Observed" choice proportions for Experiment 3, object clause version; equalities evenly divided over categories "greater than" and "less than"; *decimal point omitted*

	DM	DK	DE	DT	MK	ME	MT	KE	KT	ET
DM	-	98	98	92	79	99	85	94	82	52
DK	02	-	65	67	00	63	22	54	32	03
DE	02	35	-	46	05	46	19	31	11	00
DT	08	33	54	-	16	51	22	34	20	07
MK	21	100	95	84	-	95	84	97	86	26
ME	01	37	54	49	05	-	18	30	14	01
MT	15	78	81	78	16	82	-	66	49	14
KE	06	46	69	66	03	70	34	-	12	13
KT	18	68	89	80	14	86	51	88	-	13
ET	48	97	100	93	74	99	86	87	87	-

Both analyses thus indicate the existence of discrepancies between the compared frequencies patterns. The diagnostic subprogram routinely employed in conjunction with our goodness of fit program turned up some revealing results for the central problem of Experiment 3.

For the *interrogative* version the successive elimination procedure for word pairs contains one iteration only, in which the pair (man,koopt) is detected as the largest contributor to the overall L-value. The sum of its contributions is 22.234 corresponding to an average overprediction of .072, which indicates that the (man,koopt) cohesion in the interrogative version is less than it is in the declarative version. This decrease of the (man,koopt) cohesion is also reflected by the critical PWC (man,koopt) vs. (koopt, tas). The original asymmetry which -in terms of adjusted frequencies- was 27 to 22 "in favour of" (man,koopt), reverses in the interrogative version, where the frequencies become 21 to 29. The shift is not great, the corresponding L-contribution being no more than 3.45. Nevertheless, it reflects a small departure from both the distance interpretation with its expected recurrence of the original asymmetry, and the concatenability interpretation with its allied equality prediction. The choice frequencies for the other critical PWC, viz. (de,koopt) versus (koopt,een) lend far less support to the presumed causes of the asymmetry. The distance interpretation predicts a reversion of the original (de,koopt) vs. (koopt,een) asymmetry; the concatenability interpretation an equality. In both cases a notable L-contribution ought to be expected. However, the PWC contributes virtually nothing since the same asymmetry is observed in the interrogative version's data, 14 to 36, against 13.5 to 35.5 (adjusted) for the declarative version.

For the *object clause* version the diagnostic successive elimination procedure yields three iterations. In the first iteration (koopt,tas) is detected as a considerably overpredicted word pair (O(K,T) = .125, see Page 146). The nine PWCs in which (koopt,tas) is involved as one of the pairs contribute 98.45 (!) to the overall L of 208.37. So in spite of the reduction of superficial word distance, and of their concatenation, the words *koopt* and *tas* cohere much less in the object clause version than they do in the declarative version's data. Although (man,koopt) is deconcatenated and reversion or at least equalization of the (man,koopt) vs. (koopt,tas)-asymmetry was expected under the distance and concatenability interpretations respectively, this asymmetry is strengthened: it was 27 to 22, now it becomes 42 to 7. The associated L-contribution is 21.08. In the second iteration of the elimination procedure, after removing the contribution of the pair (koopt,tas), the pair (koopt,een) is found to be a considerably overpredicted word pair. The eight remaining PWCs having (koopt,een) as one of the members contribute 65.05 to the residual L of 109.92. The average overprediction is .122. The critical PWC (de,koopt) vs. (koopt,een) contributes 15.23 to the overall L. The original asymmetry (13.5 to 35.5) no longer or, at least, hardly exists in the object clause version's data (22.5 to 26.5). Although the figures of this second iteration do not depart too much from the patterns predicted by the distance or concatenability interpretations, this, of course, does not reverse the conclusion suggested by the first iteration and the interrogative version: it is not possible to subsume the puzzling asymmetries in the declarative version's data under generalizations relating these asymmetries to differences in distance or concatenability. Finally, for completeness, it should be mentioned that in the third step of the elimination procedure the six remaining (man,tas) pairs contribute 21.11 to the residual L of 44.87. A small underprediction of -.020 indicates that (man,tas) is somewhat more cohesive in the object clause version's data than in the declarative version's data.

In conclusion, we are able to point to significant differences between the three versions of the sample sentence as far as the choice frequencies are concerned. These differences can be localized in the pair (koopt,tas) for the interrogative version, and mainly in the pairs (koopt,een) and (koopt,tas) for the object clause version. It is surprising to see that these cases share a formal and an empirical characteristic: in every case the declarative version's word order is reversed and the cohesiveness reduced. Besides these three, the only pair whose word order also varies -under the interrogative transformation- is (de,koopt). This inversion, however, is not reflected in cohesion reduction: the nine PWCs involving (de,koopt) contribute no more than 8.59 to the overall L. We shall return to this point in the discussion section.

Syntactically unrestricted application of Luce's choice theory

The straightforward, i.e. syntactically unrestricted, application of Luce's choice theory to the interrogative version's data yielded the scale values presented in the upper diagonal cells of Table 7.6. The cells below the

Table 7.6 Unrelatedness values resulting from the linguistically unrestricted application of Luce's choice theory to the interrogative version (upper diagonal cells) and the object clause version (lower diagonal cells). For comparison, the corresponding values for the declarative version are also presented (values within parentheses)

	de	man	koopt	een	tas
de	-	.006	.334	.574	.542
		(.009)	(.287)	(.558)	(.575)
man	.014	-	.039	.478	.088
			(.025)	(.497)	(.116)
koopt	.292	.026	-	.163	.024
				(.121)	(.032)
een	.588	.532	.274	-	.007
					(.009)
tas	.435	.112	.092	.014	-

diagonal contain the scale values similarly obtained for the object clause version. In reading the matrix a caveat should be kept in mind. All subsequent Luce analyses will employ the "additive constant" adjustment discussed in the previous chapter as a successful means for the probabilization of the deterministic type 2 PWCs. The non-deterministic behaviour of these PWCs (see the 1st and 10th rows and columns of Tables 7.2 and 7.4) invites this adjustment. Accordingly, the preliminary unrestricted test of the model will be applied to all 10 word pairs. So far the preliminary tests of the Luce model have been confined to the 8 syntactically incomplete word pairs. In order to make a proper comparison possible, the scale values resulting from a similar application of the Luce model to the declarative version's data are also given in the matrix. They are presented within parentheses; the values -like those given for the interrogative and the object clause versions- have been normalized in such a way that their raw sum of squares equals unity.

The L-value obtained for the interrogative version is 30.530 (df = 36; p = .726) and the MAD-value is .026. The solution is therefore promising in terms of goodness of fit. Comparison of the upper diagonal scale values of Table 7.6 with the corresponding scale values obtained for the declarative version, confirms the result of the preliminary analysis of the previous section, viz. that the greatest difference pertains to the word pair (man, koopt). In the interrogative version's data this pair is more unrelated than it is in the declarative version's data. This appears clearly from application of "multiplicative regression"*, in accordance with the ratio character of the Luce scale values. The optimal multiplicative transformation is: interrogative scale value = .890 x declarative scale value, with an associated sum of squared relative errors of .519. On the basis of this equation the (man,koopt) value ought to be .022 whereas its empirical value is .039; the corresponding contribution to the total error sum is .180 which amounts to 38%.

The L-value for the object clause version is 56.304 (df = 36; p = .017); MAD = .041. Unlike previous cases, here Luce's choice theory yields a goodness of fit value which only borders on acceptability. The PWCs (de,koopt) vs. (man,koopt) en (koopt,een) vs. (koopt,tas) are the greatest contributors to the overall L, viz. 8.451 and 7.052, respectively, the choice frequencies

* i.e. fitting a model $\hat{Y}_i = k \times X_i$ in such a way that the sum of squared relative errors of prediction $\Sigma_i((Y_i - \hat{Y}_i)/Y_i)^2$ is minimal; elementary calculus shows that this holds for $k = \Sigma(X/Y) : \Sigma(X^2/Y^2)$.

being more extreme than those predicted on the basis of the model. Comparison of the scale values for the object clause version of Table 7.6 (lower diagonal cells) with those for the declarative version, shows clearly that the greatest differences pertain to the word pairs (koopt,een) and (koopt,tas). In accordance with the results of the preliminary analysis in the previous section it appears that these pairs are much more unrelated than in Experiment 2. Application of multiplicative regression to the scale values yields the equation: object clause value = 1.084 x declarative scale value, with an associated sum of squared relative errors of 1.057. On the basis of this equation the (koopt,een) value would have to be .131 against .274 (the "empirical" value), and the (koopt,tas) value .034 instead of .092 (the empirical value). The contributions to the total error sum are .273 and .394 respectively; a joint contribution of 63%.

Probabilistic models under different restrictions

We shall now turn to the application of Luce's choice theory to the four syntactic structures CGV, CGN, CGR and CGP for the interrogative and object clause versions' data. Although, eventually, a solution is required in which /koopt/ $\geqslant$ /man/ = /tas/ > /de/ = /een/, we shall start the analysis without imposing corresponding restrictions on the parameter space. There is no a priori reason for excluding the possibility of our requirements being met spontaneously in the parameter free case. The results of the analysis without any restrictions beyond positivity of weights are reported in Tables B1 and B2 of the Appendix. Graphical comparison of the models on the basis of the MAD-values are given in Figures 7.3 and 7.4. As mentioned in the previous subsection all subsequent Luce analyses employ the "additive constant adjustment".

The results seem to confirm the main finding of Experiment 2, viz. the "superiority" of the CGP-Luce model. The relevant figures for this model are as follows: for the interrogative version L = 59.413 (df = 40, p = .244 x 10^{-1}), MAD = .046; for the object clause version L = 71.589 (df = 40, p = .183 x 10^{-2}), MAD = .053.

Bearing in mind the conclusion reached during the preliminary step of the analysis of Experiment 3 data (viz. the attempt to predict these data from the declarative version's data), this finding is interesting. The preliminary analysis had revealed significant discrepancies between the choice frequencies for the declarative version and those for the "transformed" versions. For the interrogative version the discrepancies could be localized in the pair (man,koopt) which was judged less cohesive than it was under the

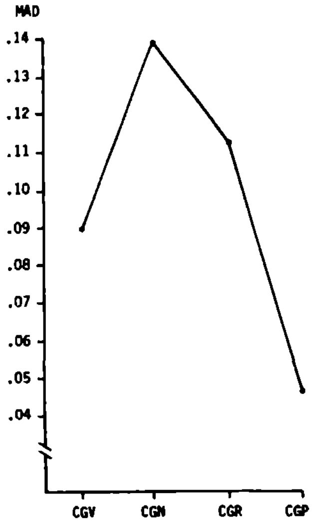


Figure 7.3

Experiment 3, interrogative version; overall MAD-values. No assumptions beyond positivity of weights (restriction a).

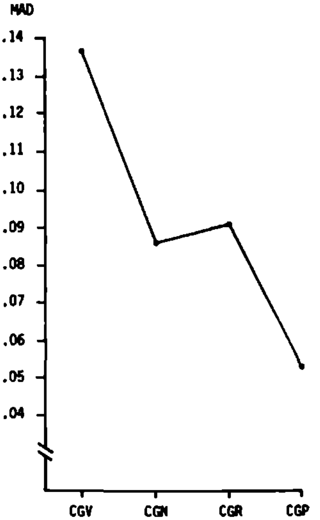


Figure 7.4

Experiment 3, object clause version; overall MAD-values. No assumptions beyond positivity of weights (restriction a).

declarative version. For the object clause version a considerable reduction in the cohesiveness of (koopt,tas) and (koopt,een) was noticed. Accordingly, shifts in the relative levels of CGV and CGN under "transformations" could be expected. As a matter of fact, Figures 7.3 and 7.4 clearly show that the fall in the (man,koopt) cohesiveness for the interrogative version favours CGV over CGN, whereas an opposite effect accompanies the decrease of the (koopt,tas) and (koopt,een) relatedness for the object clause version. But the effects are not so large as to rob the CGP-Luce model of its leading position.

For the moment, however, this conclusion must be considered as tentative. The parameter configurations obtained with CGP-Luce are far from satisfying the plausible syntactic requirements given above, as may be verified by inspection of Table 7.7.

Table 7.7 Parameter configurations obtained for the CGP-Luce model; no restrictions beyond positivity of weights

	de	man	koopt	een	tas	κ
interrogative version	.094	1.000	.150	.049	.584	.019
object clause version	.053	1.000	.074	.267	1.175	.052

Therefore, we will now turn our attention to the analyses obtained under the restriction /de/ = /een/ and /man/ = /tas/, which we shall refer to as restriction b (as we did in Chapter 6) and under the additional restriction /koopt/ $\geqslant$ /man/ = /tas/ (restriction c). The results of these analyses are presented in Tables B3, B4 and B5 of the Appendix and, as far as the MAD-values are concerned, in Figures 7.5 and 7.6. In 5 of the 8 parameter configurations obtained under restriction b, the verb outweighs the nouns which, in their turn, outweigh the articles. The exceptions are: the two CGP analyses and CGN-Luce for the object clause version. These three analyses were therefore the only ones replicated under restriction c. The results of these replications can also be found in the afore-mentioned tables and figures. The MAD-values corresponding to the CGP-analyses are plotted above the point marked CGP_c, whereas for CGN-Luce the points representing the MAD-values under the restrictions b and c coincide.

Restricting ourselves now to the least unsuccessful models satisfying /de/ = /een/ < /man/ = /tas/ $\leqslant$ /koopt/, it can be observed that for the interrogative version the superiority of the CGP-model can be maintained. For the object clause version the effects of the introduction of the restrictions b and c are larger. It is remarkable to see that CGN-Luce ($L = 167.48$, $df = 42$, $p < .0001$, $MAD = .087$) enters into competition with the symmetric model CGP-Luce ($L = 196.343$, $df = 42$, $p < .0001$, $MAD = .083$). The conclusions to be drawn from these results will largely depend on the diagnostic analyses of the least unsuccessful models. Either an adequate account of the object clause version's data will require the incorporation of CGN as the syntactic component, or it can be maintained that an underlying symmetric structure (CGP-Luce) forms part of a judgment process which is additionally co-determined by order dependent effects of an object clause "transformation".

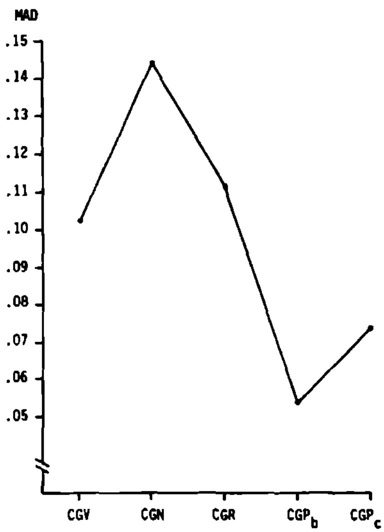


Figure 7.5

Experiment 3, interrogative version; MAD-values obtained under restrictions b and c (when not already met by the solution under restriction b).

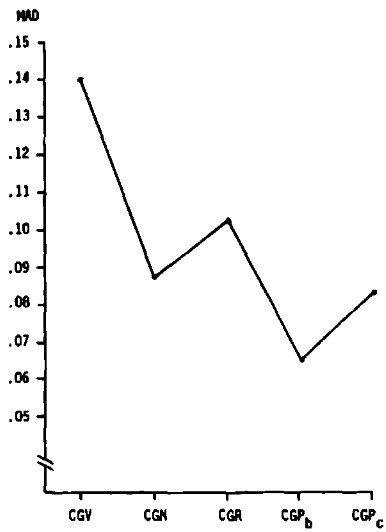


Figure 7.6

Experiment 3, object clause version: MAD-values obtained under restrictions b and c (when not already met by the solution under restriction b).

Sources of deficiencies

In this section we shall turn our attention to the diagnostics of what appeared from the analysis of the previous section to be the least unsuccessful models. For the interrogative version this was the CGP-Luce model for the object clause version CGP-Luce and CGN-Luce.

In the discussion of Chapter 6 we reported that a considerable part of the overall L-value obtained through application of CGP-Luce to the declarative version's data could be ascribed to the word pair (man,tas), the cohesion of which was underpredicted by the purely syntactic predictions. In order to prevent distortion of the model's parameter estimation and diagnosis we suggested associating a semantically interpreted (sub)additive parameter, *sem*, with this pair in order to increase the relatedness value. As there is no reason to expect that (man,tas) would behave differently in the interrogative and object clause versions we look first at the contribution of this pair to the relevant L-values. As can be seen from Table 7.8, the pair (man, tas) again yields very large contributions to the various L-values.

Table 7.8 L-contributions and "overprediction-values" of the pair (man,tas) in the least unsuccessful analyses with the Luce model

	Interrogative version	Object clause version	
	CGP-Luce	CGP-Luce	CGN-Luce
total L	144.950	196.344	167.476
contribution from the (man.tas)-PWCs	87.418	93.079	34.284
overprediction	-.167	-.175	-.093
residual L	57.532	103.265	133.192

To achieve a clear diagnosis we therefore re-analyzed the models after adjustment using the (sub)additif parameter *sem* (see Chapter 6) for the word pair (man,tas). The results are summarized in Table 7.9*. In all three cases the parameter *sem* receives a considerable negative weight, thus indicating that the adjustment does indeed substantially correct a partial deficiency in the models. This is also clearly reflected by the overall L-values which are much less than those associated with the non-adjusted applications. Furthermore, it should be noted that both the L- and MAD-values again begin supporting symmetry of structure over asymmetry: following adjustment the CGP-Luce model fits the object clause version's data better than the CGN-Luce model does.

* For a complete review of the compositions of the L-values see Tables B6 to B8 inclusive in the Appendix.

The diagnostic subroutine of the data analytic program (see Chapter 6) provided revealing results.

Table 7.9 The least unsuccessful models, reanalyzed after adjustment with the parameter *sem* for the pair (man,tas)

parameter values	interrogative version	object clause version	
	CGP-Luce	CGP-Luce	CGN-Luce
art	.114	.172	.054
noun*	1.000	1.000	1.000
verb	1.027	1.000**	2.115
κ	.031	.069	.119
sem	-.876	-.924	-1.528
L	61.308	115.314	126.469
df	41	41	41
p	$.213 \times 10^{-1}$	$.110 \times 10^{-5}$	$.270 \times 10^{-6}$
MAD	.047	.059	.071

For the *interrogative* version both the successive elimination procedure for word pairs and that for PWCs contain one iteration only. The maximally contributing word pair is (koopt,een), with an L-contribution of 25.458, corresponding to an average underprediction O(koopt,een) = -.058. After removal of the (koopt,een) contributions, the residual L amounts to 35.851 and is no longer significant (residual df = 32, p = .292). The maximally contributing PWC is (de,koopt) versus (koopt,een). The observed frequencies are 14 to 36 against 25 to 25 expected, yielding an L-contribution of 10.019. This result is not very surprising since the preliminary comparison between the declarative and interrogative versions has already brought to light how the declarative version's (de,koopt) vs. (koopt,een) asymmetry can also be discovered in the interrogative version's data. An interesting point, however,

* The noun parameter is set to unity in advance.

** The verb parameter reaches the imposed lower limit, viz. the noun-weight.

is that the decrease of the (man,koopt) cohesion which could be established in the same preliminary comparison *between* versions is not strong enough to manifest itself significantly in the *intra* interrogative version's analysis. A shift, however, can be observed. In the declarative version's analysis the (man,koopt) cohesion appeared to be theoretically underpredicted, O(man, koopt) = -.018, and the (koopt,tas) cohesion overpredicted O(koopt,tas) = +.017. This undergoes reversal in the interrogative version's analysis: O(man,koopt) = +.035 against O(koopt,tas) = -.036.

With respect to the *object clause* version we shall consider both CGP-Luce and CGN-Luce (both models under the restriction /noun/ $\lesssim$ /verb/, and adjusted with the parameters κ and *sem*). In the previous section we have seen that the strong (man,koopt) $\gtrsim_{b}$ (tas,koopt) asymmetry favours the CGN-model in such a way that it enters into competition with the CGP-Luce model. We shall now study the composition of CGN's L in detail.

The diagnostic subroutine yielded a series of eliminations of word pairs containing three steps. The maximally contributing pair is (de,koopt); its sum of contributions to the total L (126.469) amounts to 51.128. The pair is theoretically overpredicted, O(de,koopt) = .125. The greatest contributor to the residual L of 75.341 is the pair (koopt,tas) with a contribution of 32.326 (based on 8 PWCs). An O(koopt,tas) of -.102 indicates that the model seriously underpredicts the cohesiveness of this pair. This leaves a residual L of 43.015 (based on 45 - 9 - 8 = 28 PWCs) to which the seven remaining (man, tas) PWCs contribute 23.716,O(man,tas) = .060. The claim that the pair (de, koopt) and (man,koopt) cohere more strongly than (koopt,een) and (koopt,tas) reflect the very essence of the CGN incompleteness model. This claim seemed to be supported by the considerable fall in the cohesiveness of (koopt,een) and (koopt,tas) under the object clause transformation. Nevertheless, (koopt, tas) is still theoretically underpredicted, or in other words, the cohesiveness of this pair is still too strong to be accounted for by this model. At the same time *de* and *koopt* cohere much less than is predicted. Since no clear indication for non-syntactic interpretations of this counterevidence can be found, we regard these violations as crucial counterevidence against CGN-Luce. Moreover, after introduction of the parameter *sem* its fit was worse than that of CGP-Luce.

Finally, Table 7.10 gives the diagnosis of the composition of the L-value (115.319) obtained upon application of CGP-Luce to the object clause version. It clearly reflects the considerable (man,koopt) $\gtrsim$ (koopt,tas) asymmetry that crops up in the object clause version. The PWC (de,koopt) versus (koopt,een), however, is no longer on the list of contributing PWCs, since its asymmetry has vanished. The (man,koopt) $\gtrless$ (koopt,tas) asymmetry is problematic. An account of it in terms of an asymmetric underlying structure (CGN) did not seem successful. The same, however, must be said of the explanations hypothesized at the outset of this chapter in defence of the adoption of an underlying symmetric structure against the empirical asymmetry in the declarative version's data. Such explanations break down when applied to the new asymmetries in the object clause and interrogative versions.

Table 7.10 Successive elimination of PWCs and word pairs for CGP-Luce (under restriction /noun/ $\leqslant$ /verb/, adjusted with parameters κ and *sem*), applied to the object clause version's data

iteration	maximal contributor	contribution to L	over-prediction	residual L	residual df	p
a) *PWCs*						
1	MK - KT	27.737	-.357	87.573	40	.0001
2	DK - MK	19.891	.184	67.682	39	.0032
3	MK - KE	10.953	-.153	56.729	38	.0257
4	MK - MT	6.971	-.167	49.758	37	.0787
5	MT - KE	5.326	.160	44.434	36	.1580
b) *word pairs*						
1	MK	69.720	-.098	45.590	32	.056
2	KT	21.769	.068	23.821	24	.528

Rather than solving a problem, it might be said that Experiment 3 has only raised new questions. However, certain noteworthy findings should not be overlooked. Numerically at least, the CGP-Luce model maintains its leading position under the transformations studied. The new deficiencies seem mainly to be limited to those instances where the transformations lead to inversion of the declarative version's word order. In this context we can repeat our remark concluding the preliminary comparison of the data of Experiment 3 with those of Experiment 2 (see Page 182): cohesion reduction under transformations almost exclusively affects (man,koopt) in the interrogative version and (koopt,een) and (koopt,tas) in the object clause version.

7.2 CONCLUDING DISCUSSION

Experiment 3 was designed to study the determinants of the (de,koopt) versus (koopt,een) asymmetry that was established in the declarative version's data of Experiment 2. It was expected that the pattern of discrepancies between the data of Experiments 2 and 3 would enable a clear choice to be made between the alternative interpretations of the asymmetry in terms of superficial word distance or concatenability. On the basis of the analyses presented in this chapter this has turned out to be an idle hope. The pattern of discrepancies is incompatible with either of the suggested explanations. Although, strictly speaking, the results do not invalidate these explanations as far as the declarative version is concerned, their applicability to the transformed versions is falsified beyond doubt.

At the same time, however, Experiment 3 has provided us with some interesting new insights and suggestions which could form the focus of further cohesion research. Significant differences have been established between the patterns of choice frequencies for the declarative version and the "transformed" versions of Experiment 3. On this basis it could not be excluded that the frequency patterns of the latter experiment might favour other syntactic structures. This, however, was not the case. The symmetric structure CGP-Luce maintained its position as the best model on the basis of its L-value and the interpretability of its deficiencies.

In our opinion these results suggest a cohesion model for the three versions' data in which one common underlying structure should be adopted as the syntactic component. We consider CGP-Luce to be the optimal choice for this common structure, although, for reasons of improving the goodness of fit, it must first undergo several adjustments. Some of these adjustments are common to all three versions and have already been suggested and put to the test: (i) probabilization of the type 2 PWCs by the introduction of an additive constant into all word pairs' incompletenesses; (ii) the introduction of a semantically interpreted (sub)additive parameter for the word pair (man,tas).

Additional adjustments, however, for the unique aspects of the various versions seem to be indispensable in view of their significantly differing response patterns. Puzzling as the (de,koopt) vs. (koopt,een) asymmetry in the declarative version still is, since it does not seem to be subsumable under generalizations in terms of word distance or concatenability, we see that it disappears under the object clause transformation through a considerable fall in the (koopt,een) cohesion. Besides the changes of word distance and concatenability the only apparent formal change is the inversion of the word order of the pair (koopt,een). This leads to the conjecture that the cause of the asymmetry must be localized in the word order, or at least in aspects closely related to the word order of the pair (koopt,een). Something must attach to the order of the words (koopt,een) which makes them more cohesive than the equally incomplete pair (de,koopt) and which vanishes under inversion. This is a notable finding in view of the fact that, thus far, cohesion models have only interpreted the "vertical" aspects of syntactic structure, i.e. aspects which depend entirely on the dominance relation (see Chapter 2) and are invariant with respect to the "horizontal" aspects of structure. If, however, the above conjecture were in fact true, this would imply that the neglect of "horizontal" aspects of structure is not justified. Another point in this connection is the observation -by mere serendipity- that for other word pairs, viz. (man,koopt) and (koopt, tas), changes of word order covary with changes of cohesion as well. *Man* and *koopt* cohere more strongly in the declarative version than *koopt* and *man* in the interrogative version; *koopt* and *tas* cohere more strongly in the declarative version than *tas* and *koopt* in the object clause version. Order dependence of cohesion is therefore not an isolated phenomenon in the data of Experiment 2, but manifests itself far more widely than was hithero presumed. One must be prepared to find differences in cohesion for equally incomplete but differently ordered word pairs, both within and across sentences. More precisely, these are word pairs (x_1,y_1) and (y_2,x_2) where x_1 and x_2 are of the same lexical category X and y_1 and y_2 of the same lexical category Y, but with the X- and Y-representatives inversely ordered.

An interesting partial replication (Smeets, 1982) of our study is worth mentioning here. Part of this study involved the presentation of German counterparts of the declarative and object clause versions of our specimen sentence to students of German. These counterparts were *der Mann kauft einen Hut* (the man buys a hat) and (...)*dass der Mann einen Hut kauft*. It was hypothesized that for both pairs (kauft,einen) and (kauft,Hut) inversion would decrease the intuited cohesion. This, however, turned out to be significantly the case for the pair (kauft,Hut) only, but not at all for the pair (kauft,einen), in contrast with our findings for the word pair (koopt, een) in Experiment 3. The author's explanation was, that the case ending present in *einen* provides an extra clue towards the syntactic function of this article in addition to its canonical position in the declarative sentence. This extra clue which is neither present in *Hut* nor in Dutch *een*, does not disappear in the object clause version, and is therefore considered to prevent the word pair's cohesion from decreasing.

In view of the foregoing considerations, the following conclusions may be drawn: (1) for reasons of structural adequacy, future cohesion models will have to account for order dependent determinants of cohesion; (2) more specifically, the preference for the D-grammar wil be conditional on finding a solution to this order-related problem; (3) a study of the conditions under which cohesion varies with changes in word order is the most obvious course to take in furthering the research of this thesis.

At this stage there is, admittedly, insufficient empirical basis for making precise theoretical decisions as to *how*, i.e. by what formal means, and *where*, i.e. in the syntactic or non-syntactic components of the model, to account for the order dependent aspects of cohesion. Nevertheless, there is enough evidence to conclude this chapter with two suggestions concerning future research in this direction.

(i) We see in the order dependent aspects of cohesion an indication that some aspects of the left-to-right processing of the sentence play a role in the process of making cohesion judgments. In processing sentences it obviously makes a difference which word of a pair $\{x,y\}$ precedes and which follows. The syntactic commitments following an occurrence of x will generally differ from those following an occurrence of y, and the function of y in the syntactic commitments raised by x will be different from that of x in the syntactic commitments raised by y. For instance, the syntactic commitments of a preceding article exclude an immediately following verb, whereas the syntactic commitments of a transitive verb include the possibility of an immediately following article. Thus far it has been assumed that a cohesion model could neglect such factors; from now on we shall have to reckon with the possibility that such differential syntactic commitments may indeed be reflected in cohesion judgments. Perhaps this occurs in the (de,koopt) vs. (koopt,een) asymmetry noted earlier. Conjectural as these considerations are, we think that formalization of these differential syntactic commitments and the expectancies based on them might turn out to be the next best step. In this connection we think that incorporation of a left-to-right syntactic analyzer (e.g. of the augmented transition network-type; Kaplan, 1973; Woods, 1973) might prove worth considering.

(ii) A second issue for further investigation is the distinct role which the simple declarative version, or structural aspects exhibited by the declarative version, seem to play in the process of making cohesion judgments. For three of the four word pairs of the sample sentence whose word orders change under transformation, viz. for (man,koopt) in the interrogative version and (koopt,een) and (koopt,tas) in the object clause version, inversion is accompanied by a considerable decrease of cohesiveness. The only exception is the pair (de,koopt). Therefore, for some word pairs at least, a particular ordering of the words, apparently the declarative version's order, may be regarded as a salient aspect of their relationship. This "unmarked" or "canonical" order seems to predominate in the syntactic commitmens. According to this predominating canonicity in SVO languages an *agent's* privilege of occurrence is *before* a verb like *koopt*, and a *object's* privilege of occurrence is *after* the verb. When as a consequence of transformation a given word pair is presented in its non-canonical order, it becomes deprived of a salient cue to its underlying syntactic or functional relation. We think that this difference between the stimulus form of a word pair and its canonical form as reflected in the declarative version, might explain the typical cohesion reductions observed across sentences.

A final question worth considering, is whether a continuation of cohesion research along the above-mentioned lines will be likely to have consequences on the preference for D-grammars. Will we have to resort to syntactic notions that are beyond the apparatus of D-grammars or not? We are inclined to say no. One should carefully distinguish shortcomings requiring syntactic notions that can be *added to* the specifications of a given D-theory, from those which invite syntactic remedies which are alien to D-theory. Even if incorporation of a syntactic analyzer should prove necessary in accounting for the differential syntactic commitments, this in itself would not exclude D-grammars. D-grammars can be implemented in such syntactic analyzers just as well as C-grammars (Hays, 1967, p. 114 ff.). With respect to the issue of canonical order, we would have to specify for which word pairs this notion is relevant. We have already suggested that this is the case for word pairs whose declarative version's order reflects the functional relations between the words as "agent of" or "object of". These notions can also be incorporated in D-grammars as Robinson (1968) and Anderson (1971) have shown.

SUMMARY AND CONCLUSIONS

The central question of this study has been whether and, if so, how native speakers' judgments of the strength of syntactic relatedness between words and constituents of a sentence (cohesion judgments) can be used to provide empirical support toward making certain choices and decisions involved in writing a syntax. The accent has thus been on the elucidation of a methodological approach rather than on attempting to settle any linguistic dispute. In this demonstration, a comparison between the constituent (C) and dependency (D) models has been used to illustrate the argument. It is a premise of the methodological approach adopted here (see Chapter 1) that empirical phenomena such as cohesion judgments (like most data of linguistic performance) are in general to be considered very indirect reflections of syntactic structure. In dealing with these phenomena, it is the rule rather than the exception that syntactic factors are intertwined with various co-determinants of a non-syntactic or even non-linguistic nature. To the degree to which this is inherently the case, integrated models in which these factors are represented must be regarded as indispensable to a study of a given syntax's structural adequacy. This point of view links up with the "realistic approach" to syntax, as represented, for instance in Bresnan (1978).

From this point of view it is clear that a comparison of the structural adequacies of C- and D-grammars with respect to cohesion judgments could only proceed in a rather indirect fashion. The crucial issue in comparing these formalisms has been which, if either, of them best facilitates the construction of a plausible interpretation theory filling the gap between a grammatical formalism and the empirical data. The interpretation theory is thereby called upon to specify (i) which formal properties of the syntactic formalism tested are reflected in the cohesion judgments, (ii) whether and, if so, how non-syntactic determinants of a systematic or random nature affect the cohesion judgment. Of course, an advocate of such an indirect methodological approach faces an extra strong version of the well known methodological dilemma of the contrary risks of the 1st and 2nd kind. An indiscriminate resort to extraneous factors may easily protect an inadequate syntax against violations, thereby producing a "type 2 error". Nevertheless, for most types of primary or secondary verbal behaviour, the indirect approach is indispensable. To make a particular syntax solely responsible for all of the variation in the data, is to render it vulnerable to the degree

that violation can not be escaped. The risk of comitting the "type 1 error" of rejecting an adequate syntax, then becomes unacceptably high.

Taking this as our starting point, let us follow the thread running through this study and review both the promising and the unsuccessful steps that have been taken towards the development of a model for cohesion judgments.

We started (Chapter 2) with a presentation and detailed discussion of Levelt's (1974) C- and D-models for cohesion judgments. Basically, his approach involved the definition of a metric distance over the underlying C- and D-structures respectively, and the formulation of an interpretation axiom deterministically relating inter-word relatedness to metric distance in an inverse fashion. We decided against the adoption and furtherance of this approach for two main reasons. Firstly, in view of the deterministic formulation in exclusively syntactic terms the models were considered too vulnerable for exhibiting the possible virtues of the syntactic formalisms involved. Secondly, arguments were put forward (Section 2.4) for the exclusion of the entire family of distance models -to which Levelt's models belong- as far as cohesion judgments are involved.

In preparation for an alternative interpretation theory it was necessary to introduce a reformulation of the current structural descriptions given by C- and D-grammars in set-theoretical terms. This was accomplished in Chapter 3 under the inspiration of Gaifman's (1965) comparative study of the formal properties of C- and D-structures. It is argued that D-structures can be conceived of as specifying phrases and subphrases over sentences as subsets of word occurrences in a way comparable to C-grammars. Hence, both C- and D-structures can be regarded as collections of subsets of word occurrences and thus compared. The distinctive aspects of these newly conceived C- and D-structures were discussed. Whereas, unlike the D-structures, the C-structures can be of a degree higher than unity, the D-structures' collection of phrases can be and generally are, non-hierarchic as opposed to C-structures.

As an alternative to the distance models abandoned, the so-called *incompleteness models* were introduced in Chapter 4. The notion of the incompleteness of two words (or constituents) was defined as the set-theoretical complement of the words or constituents involved with respect to their smallest common constituent. On the basis of the set-theoretical inclusion relation over these incompletenesses, the word and constituent pairs of a given sentence can be partially ordered as more or less incomplete. This partial order allows an interpretation rule to derive a similarly partial order of cohesiveness over the word and constituent pairs: whenever a word pair's incompleteness is included in another word pair's incompleteness, it is predicted to be the more cohesive one. Chapter 4 concluded with an initial deterministic test of the incompleteness principle, the result of which indicated that it was a promising step forward. The stochastic nature of the data, however, invited probabilization of the incompleteness models.

Following this, Chapter 5 introduced the notion of weighted incompleteness and the possibility of incorporating the following three choice theories into the cohesion model. Incorporation of Luce's choice theory related the choice probabilities of the word pairs to the ratios of the word pairs' weighted incompletenesses. Adoption of Thurstone's choice theory made choice probabilities a normal ogive function of the differences of weighted incompletenesses. Introduction of the Restle model made the choice probabilities a function of the differential set-theoretical compositions of the word pairs' incompletenesses. These three probabilistic models were applied to four alternative syntactic representations for the structure of the sentence type: art_1-$noun_1$-*transitive verb*- art_2- $noun_2$. Sentences of this type were considered small enough to be experimentally manageable but at the same time large enough to exhibit those typical differences between C- and D-structures discussed in Chapter 3. The alternative structures were denoted as CGV (in which the verb joins the NP_2 to form a VP), CGN (in which the verb joins the NP_1 to form a "nucleus"), CGR (a reduced constituent structure in which there is neither a nucleus nor a VP) and CGP (a "pseudo constituent structure" as it is the non-hierarchical collection of substructures implied by the D-structure, of which both $\{art_1, noun_1, verb\}$ and $\{verb, art_2, noun_2\}$ are members). Among these, CGV and CGN are of degree 2 and accordingly exclusively C-structures. CGP is non-hierarchical and is exclusively a D-structure. CGR is a C-structure of degree 1, and hence -according to the formal comparison of C- and D-grammars- an equivalent D-grammar yielding the same set-theoretical structure can be constructed for it.

Application of the afore-mentioned three choice models to these four alternative syntactic structures yielded twelve models for the cohesion judgments for any sentence of the given type. Evidence gathered from two experiments reported in Chapter 6, provided reasons for the elimination of several facets of this 3 x 4 collection of models. With respect to the syntactic structures it was possible to eliminate the asymmetric models CGV and CGN (with the verb clustering either with NP_1 or with NP_2) in favour of the symmetric ones: CGR and CGP. In other words, the set of competing models could be narrowed down to a subset of models whose syntactic components are either of the dependency type or of a trivial constituent type that can be imitated by a D-grammar.

As far as the choice theories are concerned, the models could be narrowed down to the subclass of Luce models. The Restle models had to be eliminated because they give rise to inappropriately large deterministic components, i.e. subsets of pairwise comparisons for which deterministic predictions were derived. As for the Thurstone approach there were two critical arguments. Firstly, even its unrestricted and thus most lenient application resulted in a significant mismatch between predictions and data. More serious, from the viewpoint of the methodological approach adopted in this study, is the second argument: the optimal application under syntactic restrictions, viz. CGR-Thurstone, yielded a pattern of deficiencies which was hardly interpretable in terms of nonsyntactic factors. The Luce approach was not without inadequacies either. In its original form, as formulated in Chapter 5, deterministic predictions were derived for all pairwise comparisons in which one complete and one incomplete word pair are compared. The evidence, reported in Chapter 6, indicated that such a deterministic component was not justified. A minor adjustment, however, in the form of the introduction of an additive constant to all the word pairs' incompletenesses, sufficed to cure the Luce models on this point. There seem to be no comparably simple measures for remedying the other choice models' shortcomings.

Within the remaining subset of models, CGP-Luce (the ratio rule applied to the set-theoretically formulated D-grammar) was found the best fitting, or rather, least unsuccessful model At this point in the study it was tempting to make a decision in favour of the D-grammar, but the overall goodness of fit of the model was still disappointing. However, detailed inspection of the goodness of fit values made it possible to trace this poor fit to two sources of deficiencies. Firstly, the pattern of discrepancies between predictions and data strongly suggested the influence of a semantic factor, the neglect of which not only seemed to worsen the goodness of fit, but also to distort the parameter estimation. A semantically adjusted version of CGP-Luce was proposed and its test revealed a considerable improvement of the model in both respects. Secondly, there was a puzzling asymmetry in the data, which contradicted the syntactic symmetry implied by CGP-Luce, but which was not so strong as to justify a choice in favour of an asymmetric syntactic structure. The asymmetry concerned the word pairs (art_1,verb) and (verb,art_2) with the former pair being less cohesive than the latter. For this asymmetry plausible interpretations in terms of superficial word distance or "concatenability" came up for consideration. These interpretations were, moreover, consistent with another, albeit non-significant asymmetry in the data, viz. ($noun_1$,verb) versus (verb,$noun_2$) with the first word pair being more cohesive than the second. The methodological dilemma of contrary risks thus cropped up again. Would word distance or concatenability provide justifiable explanations for the poor fit of a model in which a D-grammar was rightly incorporated, or would they merely furnish ad hoc means for shielding an inadequate syntax against violation?

The experiment reported in Chapter 7 was intended to obtain a partial empirical answer to these questions. To this end the model sentence of the second experiment of Chapter 6 (*de man koopt een tas*; Eng.: the man buys a bag) was "transformed" into (a) its interrogative version and (b) its object clause version. Use of these two versions should give rise to different response patterns under the superficial word distance and concatenability interpretations. The experiment yielded unexpected results. In terms of relative goodness of fit, again CGP appeared to be the optimal syntactic option, both for the interrogative and for the object clause version's data. But again, the associated goodness of fit measures were significantly high, thus calling for detailed inspection of the discrepancies between model and data. Surprisingly, the choice frequencies for the transformed versions significantly differing from each other as well as from the declarative version's data, fitted neither the word distance pattern nor the concatenability pattern. In other words, they failed to give the CGP-Luce model the indirect support that was to be expected on the basis of the experiments reported in Chapter 6.

Instead, the experimental data reported in Chapter 7 strongly suggest that the cause of the puzzling asymmetries should be localized in differences of word order or in factors that go with these differences in word order. Order dependence of cohesion didn't appear to be limited to the obstinate asymmetry, but manifested itself more generally. More instances were found where (equally incomplete) word pairs varied in perceived cohesion along with variation of their word order, sometimes in a very systematic way. In those cases, for instance, where transformations induce inversion of the simple declarative version's word order, a significant fall in the intuited degree of cohesion could generally be observed. The tentative interpretation drawn from this was that for some word pairs at least, the declarative version's word order should be considered as the "canonical" one; it is a salient aspect of the syntactic relation and as such it is an important determinant of the perceived cohesion.

In conformity with the above facts it was argued that, thus far, cohesion models have unjustly neglected these order related aspects of cohesion. Likewise, it is clear that the preference for the D-grammar should not be made without qualification. The D-grammar should be paired with an interpretation theory capable of dealing with the word order dependent aspects of cohesion judgments. Further inquiry into the conditions under which variation of word order co-varies with intuited cohesion will be necessary before this factor can be assigned its appropriate place in the theory.

In this thesis it has become clear that cohesion judgments are indirect manifestations of syntactic structures. They reflect both random error and factors of a non-syntactic nature as well. An operationalist, therefore, would characterize this kind of data as both "unreliable" and "invalid". He will consequently look out for "better" operationalizations in the same vein als Labov's "What is a linguistic fact?" represented a reaction to the unreliability and invalidity of acceptability judgments. Perhaps the methodological distance between syntax and empirical data is greater for cohesion judgments than it is for other kinds of empirical data such as acceptability judgments or primary speech phenomena. Nevertheless, it is our opinion that this is a difference of degree rather than of kind. Accordingly we do not think that the operationalistic reaction is an essential solution to the problem of methodological distance. For this reason we agree with Chomsky's (1965, p.19) statement "...there is no reason to expect that reliable operational criteria for the deeper and more important notions of linguistics (such as "grammaticalness" and "paraphrase") will ever be forthcoming". For the same reason, however, we cannot neglect Labov's criticism of much linguistic practice. The indirect relation between syntax and empirical data holds for linguists' introspective judgments as well. These judgments are still useful and should therefore be utilized as vehicles for constructing theories, but they ought to be employed with circumspection for testing purposes. We believe that the set of privileged investigators with direct access to the "true" aspects of noisy data is empty, and consequently we think that the intersection of this set and the set of linguists is empty as well.

The "gap between the languages of theory and research" as Blalock (1971) characterizes the same problem in social methodology, should be bridged by adoption of an auxiliary theory (Blalock's term) or interpretation theory (Levelt's term), relating "general theory" to empirical phenomena. Without the guidance of such an interpretation theory the methodological status of, for instance, cohesion data is indeterminate, and resort to them may lead to quasi arguments of the sort depicted at the very beginning of this study. It was only under the guidance of an interpretation theory that seeming a-symmetries in the data could be dismantled as counterevidence to hypothesized syntactic symmetrie. In this vein, introspective evidence can be used for testing linguistic theories, as long as it is explicitly paired with an "auxiliary" interpretation theory. This brings us back to Bresnan's (1978) "realistic approach". In testing a linguistic theory it is generally inevitable to construct an integrated performance theory, while bearing in mind the important question of whether or not this linguistic theory is an indispensable component of such an integrated theory.

SAMENVATTING EN CONCLUSIES

De centrale vraag van dit proefschrift is of en, zo ja, hoe oordelen van "moedertaalsprekers" betreffende de sterkte van syntactische relaties tussen woorden en constituenten binnen een zin (cohesie-oordelen) gebruikt kunnen worden als empirische ondersteuning bij keuzen en beslissingen die zich voordoen bij het schrijven van een syntaxis. Het accent ligt daarbij op de demonstratie van een methodologische benadering en niet op het streven een definitieve oplossing te vinden voor een linguïstisch probleem. Bij deze demonstratie wordt een vergelijking van het constituentenmodel (C) met het dependentiemodel (D) gebruikt ter illustratie van het betoog. Uitgangspunt van de hier gekozen methodologische benadering (zie Hoofdstuk 1) is, dat empirische verschijnselen zoals cohesie-oordelen (evenals de meeste taalgebruiksgegevens) in het algemeen moeten worden beschouwd als zeer indirecte manifestaties van syntactische structuur. Voor deze verschijnselen is het eerder regel dan uitzondering dat syntactische factoren verstrengeld zijn met diverse factoren van niet-syntactische of zelfs niet-linguïstische aard. In de mate waarin dit inherent het geval is, worden geïntegreerde modellen waarin deze factoren vertegenwoordigd zijn onmisbaar geacht bij de bestudering van de structurele adequaatheid van een gegeven syntaxis. Dit standpunt sluit aan bij de "realistische benadering" tot syntaxis zoals o.a. vertegenwoordigd door Bresnan (1978).

Vanuit dit standpunt beschouwd is het duidelijk dat een vergelijking van de structurele adequaatheid van C- en D-grammatica's aan de hand van cohesie-oordelen nauwelijks anders dan op indirecte wijze kan geschieden. Het kardinale punt bij deze vergelijking is, welk van beide formalismen, indien al een van beide, zich het best leent voor de constructie van een plausibele interpretatietheorie, mediërend tussen grammaticaal formalisme enerzijds en de empirische gegevens anderzijds. Taak van de interpretatietheorie is daarbij te specificeren (i) welke formele eigenschappen van het te toetsen syntactische formalisme in het cohesie-oordeel worden gereflecteerd, (ii) of en, zo ja, hoe niet-syntactische determinanten van systematische of toevallige aard het cohesie-oordeel mede beïnvloeden. Vanzelfsprekend ziet een voorstander van een dergelijke indirecte methodologische benadering zich geconfronteerd met een extra sterke versie van het welkbekende methodologische dilemma van de tegenovergestelde risico's van de eerste en de tweede soort. Een lichtvaardig beroep op externe factoren kan

een inadequate syntaxis al te gemakkelijk beschermen tegen schendingen en daarbij leiden tot een "fout van de tweede soort". Niettemin is voor de meeste vormen van primair of secundair taalgedrag een indirecte benadering onvermijdbaar; door een bepaalde syntaxis exclusief verantwoordelijk te maken voor alle variatie in de gegevens maakt men deze syntaxis zo kwetsbaar dat schending onafwendbaar is Het risico een "fout van de eerste soort" te maken, te weten het verwerpen van een adequate syntaxis, wordt dan onaanvaarbaar hoog.

Laten we vanuit dit gezichtspunt de hoofdlijn van deze studie volgen en een overzicht geven van zowel de minder geslaagde als de meer belovende stappen die genomen zijn bij het ontwikkelen van een model voor cohesieoordelen.

Hoofdstuk 2 begint met de presentatie en een gedetailleerde bespreking van Levelts (1974) constituenten- en dependentiemodel voor cohesie-oordelen. In essentie behelst diens benadering de definitie van een afstandsmetriek over de aan zinnen ten grondslag liggende C- respectievelijk D-structuren, en de formulering van een interpretatie-axioma dat op deterministische wijze de relatiesterkte tussen woorden relateert aan de afstand tussen die woorden. Om twee redenen wordt ervan afgezien deze benadering over te nemen en verder te ontwikkelen. Ten eerste worden Levelts modellen, gezien de deterministische formulering in exclusief syntactische termen, te kwetsbaar geacht om de mogelijke deugdzaamheid van de betrokken syntactische formalismen te kunnen aantonen. Ten tweede worden er argumenten naar voren gebracht (Paragraaf 2.4) voor het uitsluiten van de volledige familie van afstandsmodellen ter representatie van cohesie-oordelen, waartoe Levelts modellen behoren.

Ter voorbereiding van een alternatieve interpretatietheorie is het wenselijk een verzamelingstheoretische herformulering te geven van de gangbare structurele beschrijvingen van constituenten- en dependentiegrammatica's. Dit geschiedt in Hoofdstuk 3, geïnspireerd door Gaifmans (1965) vergelijkende bestudering van de formele eigenschappen van C- en D-structuren. D-structuren, zo wordt betoogd, kunnen geledingen en ondergeledingen van zinnen in zinsleden specificeren op een wijze die vergelijkbaar is met de manier waarop dit in C-grammatica's geschiedt. Dit houdt in dat zowel C- als D-structuren kunnen worden beschouwd als collecties van deelverzamelingen van woordvoorkomens en als zodanig vergelijkbaar zijn. De aspecten waarin de aldus opgevatte C- en D-structuren verschillen worden besproken. In één interpretatie zijn de zinsgeledingen van D-structuren, anders dan die van C-structuren, beperkt tot graad 1 (§ 3.3); in een alternatieve interpretatie wijken zij van die van C-structuren af, doordat ze in het algemeen niet-hiërarchisch zijn.

Als alternatief voor de verworpen afstandsmodellen worden in Hoofdstuk 4 de zogeheten *incompleetheidsmodellen* geïntroduceerd. De *incompleetheid* van een woord- of constituentenpaar wordt gedefinieerd als het verzamelingstheoretische complement van de betreffende woorden of constituenten met betrekking tot hun kleinste gemeenschappelijke constituent. Op basis van de verzamelingstheoretische inclusierelatie over deze incompleetheden kunnen de woord- en constituentenparen van een gegeven zin partieel worden geordend als meer of minder incompleet. Deze partiële volgorde maakt het mogelijk, middels een interpretatieregel, de cohesie van de woord- en constituentenparen op een eveneens partiële wijze te ordenen: wanneer de incompleetheid van een woordpaar geïncludeerd is in die van een ander woordpaar wordt voor het eerste paar een grotere cohesie voorspeld dan voor het tweede. Hoofdstuk 4 wordt besloten met een eerste deterministische toetsing van het incompleetheidsbeginsel, dat op grond van het resultaat als een veelbelovende stap voorwaarts kan worden beschouwd. De stochastische aard van de gegevens, nodigt er evenwel toe uit de incompleetheidsmodellen een probabilistische gedaante te geven.

Deze herformulering geschiedt in Hoofdstuk 5. Eerst wordt de notie "gewogen incompleetheid" geïntroduceerd. Verder wordt besloten tot het beproeven van de mogelijkheid tot incorporatie van een drietal psychologische keuzetheorieën in het cohesiemodel. Incorporatie van de keuzetheorie van Luce houdt in dat bij paarsgewijze vergelijking van woordparen, de keuzekansen worden gerelateerd aan de verhoudingen van de gewogen incompleetheden van de woordparen. Met de introductie van Thurstone's keuzetheorie worden de keuzekansen beschouwd als een cumulatieve normale verdelingsfunctie van de verschillen tussen de gewogen incompleetheden. Verder wordt voorgesteld het model van Restle te beproeven waarbij de keuzekansen gezien worden als een functie van de differentiële verzamelingstheoretische samenstelling van de incompleetheden van de betreffende woordparen. Besloten wordt deze drie probabilistische modellen toe te passen op vier alternatieve syntactische representaties van de structuur van zinnen van het type: *lidwoord*$_1$*-zelfstandig naamwoord*$_1$*-overgankelijk werkwoord-lidwoord*$_2$*-zelfstandig naamwoord*$_2$. Zinnen van dit type zijn klein genoeg om experimenteel hanteerbaar te zijn en tegelijkertijd groot genoeg om die typische verschillen te kunnen vertonen die er overeenkomstig de vergelijkende bestudering in Hoofdstuk 3 bestaan tussen C- en D-structuren. De alternatieve structuren worden aangeduid als CGV (waarin het werkwoord met de NP_2 een VP vormt), CGN (waarin het werkwoord samen met de NP_1 een "nucleus" vormt), CGR (een "gereduceerde" constituentenstructuur waarin noch een nucleus noch een VP voorkomt) en CGP (een "pseudo constituentenstructuur", zijnde de niet-hierarchische collectie van substructuren geïmpliceerd in de D-structuur, waarvan zowel {lidwoord_1,zelfstandig naamwoord_1,werkwoord} als {werkwoord,lidwoord_2,zelfstandig naamwoord_2} deel uitmaken). Van deze structuren zijn CGV en CGN van graad 2 en bijgevolg exclusief van het constituententype. CGP is non-hierarchisch en exclusief een D-structuur. CGR is een C-structuur van graad 1 zodat -blijkens de formele vergelijking van C- en D-grammatica's- een equivalente D-grammatica met identieke verzamelingstheoretische structuur ervoor kan worden geconstrueerd.

Toepassing van de genoemde keuzetheorieën op deze vier alternatieve syntactische structuren resulteert in twaalf modellen voor cohesie-oordelen betrekking hebbend op elke zin van het gegeven type. Evidentie, verkregen uit twee experimenten, gerapporteerd in Hoofdstuk 6, geeft aanleiding tot de eliminatie van verscheidene facetten uit deze 3 x 4 verzameling van modellen. Met betrekking tot de syntactische structuren is het mogelijk de asymmetrische modellen CGV en CGN (waarin het werkwoord met NP_2 respectievelijk NP_1 clustert) te elimineren ten gunste van de symmetrische modellen CGR en CGP. Met andere woorden, de verzameling van concurrerende modellen kan worden ingeperkt tot een deelverzameling van modellen waarvan de syntactische componenten hetzij van het dependentietype zijn, of van een "eerste graads" constituententype, dat door een D-grammatica kan worden geïmiteerd.

Wat betreft de keuzetheorieen kan de verzameling van modellen worden ingeperkt tot de subklasse der Luce-modellen. De Restle-modellen moeten worden verworpen op grond van hun empirisch onhoudbaar grote deterministische component, d.w.z. een deelverzameling van paarsgewijze vergelijkingen waarvoor deterministische predicties worden afgeleid. Met betrekking tot de Thurstone benadering zijn er twee kritieke argumenten. Ten eerste leidt zelfs de onvoorwaardelijke toepassing van deze keuzetheorie, d.w.z. een toepassing zonder de restricties van een linguïstische interpretatietheorie, tot een significante discrepantie tussen voorspellingen en gegevens.

Zwaarder wegend, vanuit het standpunt van de methodologische benadering in dit proefschrift, is het tweede argument: de optimale toepassing onder syntactische restricties, nl. CGR-Thurstone, leidt tot een patroon van schendingen dat nauwelijks interpreteerbaar is via een beroep op non-syntactische factoren.

Ook de Luce-modellen zijn niet vrij van onvolkomenheden. Zoals geformuleerd in Hoofdstuk 5, leiden zij tot deterministische predicties voor alle paarsgewijze vergelijkingen van één compleet en één incompleet woordpaar. Uit de evidentie, gerapporteerd in Hoofdstuk 6, blijkt dat een dergelijke deterministische component niet gerechtvaardigd is. Een kleine aanpassing, evenwel, in de vorm van de introductie van een additieve constante in de incompleetheden van alle woordparen, voldoet teneinde de Luce-modellen op dit punt te verbeteren. Er lijken geen vergelijkbaar eenvoudige maatregelen te bestaan om het Restle-model op dit punt te cureren.

Binnen de resterende deelverzameling van modellen blijkt CGP-Luce (Luce's keuzetheorie toegepast op de verzamelingstheoretisch geformuleerde D-structuur) het best passende, of liever, minst ontoereikende model. In deze fase van het onderzoek is het verleidelijk te kiezen voor de D-grammatica; de globale "goodness of fit" van het CGP-Luce model, evenwel, is onbevredigend. Gedetailleerde analyse van de "goodness of fit"-maten maakt het mogelijk de discrepanties tussen model en data te herleiden tot twee bronnen van schendingen. Ten eerste suggereren deze discrepanties duidelijk de invloed van een semantische factor. Het blijkt dat de veronachtzaming van deze factor behalve verslechtering van de "goodness of fit" ook distorsie van de parameterschatting ten gevolge heeft. Een op dit punt gecorrigeerde versie van het CGP-Luce model wordt voorgesteld en toepassing ervan leidt tot aanmerkelijke verbetering in beide opzichten. Ten tweede doet zich een problematische asymmetrie in de gegevens voor die strijdig is met de symmetrie van syntactische structuur welke door het CGP-Luce model wordt geïmpliceerd. De asymmetrie is evenwel niet zo sterk dat zij een keuze voor een asymmetrische zinsstructuur rechtvaardigt. De asymmetrie betreft de paren (lidwoord_1,werkwoord) en (werkwoord,lidwoord_2), waarbij de cohesie binnen het tweede paar sterker blijkt dan die binnen het eerste. Als mogelijke verklaring voor deze asymmetrie wordt aangevoerd dat de betreffende woordparen verschillen in het opzicht van woordafstand en "concateneerbaarheid". Deze interpretaties zijn bovendien in overeenstemming met een andere, zij het niet significante, asymmetrie in de gegevens, nl.

(zelfstandig naamwoord$_1$,werkwoord) versus (werkwoord,zelfstandig naamwoord$_2$), waarbij het eerste woordpaar een grotere cohesie vertoont dan het tweede. Het methodologische dilemma van tegenovergestelde risico's doet zich hier weer voor. Zijn woordafstand of concateneerbaarheid te rechtvaardigen als verklaringen voor de empirische inadequaatheid van een model waarin een D-grammatica terecht is opgenomen, of zijn zij wellicht de ad hoc middelen om een inadequate syntaxis voor schending te behoeden?

Het in Hoofdstuk 7 gerapporteerde experiment beoogt een partieel empirisch antwoord te geven op deze vragen. Hiertoe wordt de zin uit het tweede experiment van Hoofdstuk 6 (*de man koopt een tas*) na "transformatie" tot vraagzin respectievelijk lijdend voorwerpszin aan proefpersonen ter beoordeling aangeboden. Deze twee versies leiden onder de interpretaties van woordafstand en concateneerbaarheid tot verschillende responsiepatronen. Het experiment levert onverwachte resultaten op. In het opzicht van relatieve "goodness of fit" blijkt de CGP wederom de optimale syntactische keuze, zowel voor de gegevens van de vraagzin als voor die van de lijdend voorwerpszin. Wederom, evenwel, vragen de significant hoge "goodness of fit" maten om een gedetailleerd onderzoek naar de afwijkingen tussen het model en de gegevens. De keuzefrequenties verkregen voor de getransformeerde versies verschillen zowel ten opzichte van elkaar als ten opzichte van die welke verkregen zijn voor bovengenoemde stellende zin. Bovendien passen zij niet in de patronen, voorspeld onder de interpretaties van woordafstand of concateneerbaarheid. Met andere woorden, zij verlenen het CGP-Luce model niet de indirecte steun die verwacht kon worden op grond van de experimenten in Hoofdstuk 6.

In plaats daarvan suggereren de experimentele gegevens, gerapporteerd in Hoofdstuk 7, dat de oorzaak van de onverklaarde asymmetrie, gesignaleerd in Hoofdstuk 6, gezocht moet worden in woordordeverschillen binnen de betrokken woordparen of in factoren die hiermee nauw samenhangen. De samenhang tussen cohesie en woordvolgorde blijkt niet beperkt tot de reeds aangeduide, onverklaarde asymmetrie, maar manifesteert zich meer algemeen. Er worden meer voorbeelden gevonden van cohesieverschillen tussen woordparen van gelijke syntactische incompleetheid, doch met tegenovergestelde woordorde. Deze voorbeelden kenmerken zich bovendien door een duidelijke systematiek. In die gevallen, bijvoorbeeld, waarin transformaties omkering bewerkstelligen van de woordvolgorde in de stellende hoofdzin, wordt in het algemeen een significante daling in de intuïtief ervaren relatiesterkte geobserveerd. De voorlopige interpretatie hiervan is, dat voor sommige woordparen de volgorde in de stellende zin als "canonisch" moet worden opgevat; deze volgorde fungeert als een saillant aspect van de syntactische relatie en is als zodanig een belangrijke determinant van de gepercipieerde cohesie.

Uit genoemde overwegingen kan worden geconcludeerd dat cohesiemodellen tot nog toe ten onrechte de afhankelijkheid van het cohesie-oordeel ten opzichte van aspecten van woordvolgorde hebben veronachtzaamd. Eveneens is het duidelijk dat de preferentie voor de D-grammatica niet zonder reserve kan worden geuit. De D-grammatica zou gecombineerd moeten worden met een interpretatietheorie, die in staat is de afhankelijkheid van het cohesie-oordeel ten opzichte van woordvolgorde het hoofd te bieden. Verder onderzoek naar de condities waaronder het cohesie-oordeel covarieert met variatie van woordvolgorde zal nodig zijn alvorens woordvolgorde een verantwoorde plaats in de theorie kan worden toegekend.

Uit dit onderzoek komt duidelijk naar voren dat cohesie-oordelen indirecte manifestaties zijn van syntactische structuren. Ze weerspiegelen zowel random error als factoren van niet-syntactische aard. Een operationalist zou derhalve geneigd kunnen zijn dit soort gegevens als "onbetrouwbaar" en "invalide" te beschouwen. Hij zal bijgevolg uitzien naar "betere" operationalisaties in dezelfde zin als waarin Labovs "What is a linguistic fact" (1975) reageert op de onbetrouwbaarheid en invaliditeit van acceptabiliteitsoordelen. Wellicht is de methodologische kloof tussen syntaxis en empirische gegevens bij cohesie-oordelen groter dan bij andere soorten van empirische gegevens zoals acceptabiliteitsoordelen of primaire taalgebruiksgegevens. In onze opvatting, evenwel, is dit slechts een gradueel, niet een essentieel verschil. Dienovereenkomstig wordt hier een operationalistische reactie niet gezien als een wezenlijke overbrugging van de genoemde methodologische kloof. Om deze reden stemmen we in met Chomskys (1965, p. 19) uitspraak dat er geen reden is om te verwachten dat er ooit betrouwbare operationele criteria voor de meer fundamentele en belangrijke linguïstische begrippen zoals "grammaticaliteit" en "paraphrase" zullen komen. Om dezelfde reden echter kan Labovs kritiek op veel linguïstische praktijk niet worden veronachtzaamd. De methodologische afstand tussen syntaxis en empirische gegevens doet zich ook voor bij de introspectieve oordelen van linguïsten. Deze oordelen blijven nuttig en het aanwenden waard als werktuigen bij het ontwikkelen van theorieën, maar zij dienen met grote omzichtigheid te worden omgeven waar het toetsingsdoeleinden betreft. Wij geloven dat de verzameling van geprivilegieerde onderzoekers met directe toegang tot de "ware" aspecten van "ruisende" gegevens leeg is en dat dit derhalve ook geldt voor de doorsnee van deze verzameling met die van de linguïsten.

De "kloof tussen de talen van theorie en onderzoek" -zo karakteriseert Blalock (1971) hetzelfde probleem in de methodologie der sociale wetenschappen- zou moeten worden overbrugd middels de incorporatie van een hulptheorie (in Blalocks terminologie) of interpretatietheorie (à la Levelt), die de "algemene theorie" moet relateren aan empirische verschijnselen. Zonder het kompas van een dergelijke interpretatietheorie is de methodologische status van, bijvoorbeeld, cohesiedata onbeslist en kan een beroep daarop leiden tot schijnargumenten van het soort dat in de aanhef van dit proefschrift is beschreven. Slechts op het kompas van een interpretatietheorie konden asymmetrieën in de gegevens worden onderkend als schijnargumenten tegen veronderstelde syntactische symmetrie.

In deze zin kan introspectieve evidentie worden gebruikt voor het toetsen van een linguïstische theorie mits deze expliciet ondersteund wordt door een interpretatietheorie. Dit voert ons terug tot Bresnans (1978) "realistische benadering". Het toetsen van een linguïstische theorie komt vaak onvermijdelijk neer op het construeren van een geïntegreerde performancetheorie waarbij als cruciale vraag moet worden gesteld, of deze linguïstische theorie daarin een onmisbare component is.

APPENDIX

Table : A1				
corresponding to Section: 6.2				
Experiment : 1				
Sentence	de man koopt het boek			
Choice model:	Luce			
Restrictions:	none except positivity of weights			
Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.000[3]	.097	.000	.017
$noun_1$	1.000[4]	1.000	1.000	1.000
verb	.442	3.276	4.093	.316
art_2	.022	.000	.000	.028
$noun_2$	.053	.000	.368	.134
MAD-values				
overall	.121	.165	.154	.099
det. PWCs[1]	.054	.054	.054	.054
prob. PWCs[2]	.154	.226	.210	.124
L-values	220.68	352.65	316.65	149.35
df	25	25	25	25
p	5			
z-transf.[6]	14.01	19.56	18.16	10.28

Footnotes pertaining to Tables A1 - A15 and B1 - B5

1) deterministic PWCs

2) probabilistic PWCs

3) in this and all other .000 cases the parameter value reaches the imposed lower limit

4) in all Luce and Restle analyses the $/noun_1/$ parameter is put to unity arbitrarily

5) p-values have been omitted whenever they are less than .0001

6) i.e. the well know transformation

$$z = \sqrt{2 \times \chi^2} - \sqrt{2 \times df - 1},$$

with L substituted for χ^2 ; the z-transformed values should facilitate comparisons of L-values for different df's

Table : A2
corresponding to Section: 6.2
Experiment : 1
Sentence : de man koopt het boek
Choice model: Thurstone
Restrictions: none except positivity of weights

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.000	.000	.000	.908
$noun_1$	1.165	1.124	1.060	2.272
verb	1.349	1.568	1.783	.000
art_2	.225	.000	.395	.966
$noun_2$	.671	.104	.808	1.411
MAD-values				
overall	.124	.144	.105	.094
det. PWCs	-	-	-	-
prob. PWCs	.124	.144	.105	.094
L-values	248.54	334.96	203.92	178.02
df	42	42	42	42
p				
z-transf.	13.18	16.77	11.08	9.76

Table : A3
corresponding to Section: 6.2
Experiment : 1
Sentence : de man koopt het boek
Choice model: Restle
Restrictions: none except positivity of weights

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.096	.109	.061	.064
$noun_1$	1.000	1.000	1.000	1.000
verb	.653	2.730	3.475	.398
art_2	.070	.009	.308	.083
$noun_2$	.154	.019	.528	.204
MAD-values				
overall	.138	.163	.121	.105
det. PWCs	.107	.161	.054	.066
prob. PWCs	.173	.166	.158	.149
L-values	172.26	178.03	228.22	140.97
df	17	17	25	17
p				
z.transf.	12.82	13.12	14.36	11.05

Table : A4
corresponding to Section: 6.2
Experiment : 1
Sentence : de man koopt het boek
Choice model: Luce
Restrictions: $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.194	.088	.000	.065
$noun_1$	1.000	1.000	1.000	1.000
verb	4.249	3.585	6.266	.632
art_2				
$noun_2$				
MAD-values				
overall	.158	.170	.160	.126
det. PWCs	.054	.054	.054	.054
prob. PWCs	.216	.234	.218	.166
L-values	347.11	387.18	325.75	229.43
df	27	27	27	27
p				
z-transf.	19.07	20.55	16.64	14.14

Table : A5
corresponding to Section: 6.2
Experiment : 1
Sentence : de man koopt het boek
Choice model: Thurstone
Restrictions: $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.010	.000	.049	.879
$noun_1$	.856	.626	.874	1.710
verb	1.385	1.265	1.702	.000
art_2				
$noun_2$				
MAD-values				
overall	.132	.166	.117	.117
det. PWCs	-	-	-	-
prob. PWCs	.132	.166	.117	.117
L-values	291.70	425.16	253.60	255.24
df	44	44	44	44
p				
z.transf.	14.83	19.83	13.20	13.27

Table : A6				
corresponding to Section: 6.2				
Experiment : 1				
Sentence : de man koopt het boek				
Choice model: Restle				
Restrictions: $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$				
Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.242	.197	.244	.184
$noun_1$	1.000	1.000	1.000	1.000
verb	2.742	3.711	4.639	.685
art_2				
$noun_2$				
MAD-values				
overall	.151	.172	.142	.125
det. PWCs	.107	.161	.054	.065
prob. PWCs	.201	.186	.190	.193
L-values	215.50	217.21	287.49	206.81
df	19	19	27	19
p				
z-transf.	14.68	14.76	16.70	14.25

Table : A7 (Residual analyses)				
corresponding to Section: 6.2 (Page 143)				
Experiment : 1				
Sentence : de man koopt het boek				
Choice mod:	Thurstone		Luce	Restle
Restrict.:	/verb/ ≥ /art/	/verb/ ≥ /noun/	/verb/ ≥ /noun/	/verb/ ≥ /noun/
Syntactic options	CGP	CGP	CGP	CGP
Weights				
art_1	.503	.204	.073	.186
$noun_1$	1.381	.976	1.000	1.000
verb	.503	.976	1.000	1.000
art_2				
$noun_2$				
MAD-values				
overall	.124	.144	.127	.125
det. PWCs	-	-	.054	.065
prob. PWCs	.124	.144	.167	.194
L-values	276.42	327.75	231.52	207.85
df	44	44	28	20
p				
z.transf.	14.19	16.28	14.10	14.14

Table : A8

corresponding to Section: 6.3

Experiment : 2

Sentence : de man koopt een tas

Choice model: Luce

Restrictions: none except positivity of weights

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.000	.167	.000	.071
$noun_1$	1.000	1.000	1.000	1.000
verb	.773	2.186	3.304	.348
art_2	.079	.000	.000	.082
$noun_2$	.263	.434	.745	.487
MAD-values				
overall	.116	.139	.122	.072
det. PWCs	.087	.087	.087	.087
prob. PWCs	.132	.167	.141	.064
L-values	212.68	323.68	227.95	67.14
df	25	25	25	25
p				
z-transf.	13.62	18.44	14.35	4.59

Table : A9

corresponding to Section: 6.3

Experiment : 2

Sentence : de man koopt een tas

Choice model: Thurstone

Restrictions: none except positivity of weights

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.000	.000	.000	.672
$noun_1$	.633	.747	.704	1.454
verb	.989	1.062	1.255	.000
art_2	.132	.000	.176	.619
$noun_2$	.654	.325	.694	1.148
MAD-values				
overall	.105	.124	.075	.077
det. PWCs	-	-	-	-
prob. PWCs	.105	.124	.075	.077
L-values	254.47	334.55	155.36	156.68
df	42	42	42	42
p				
z.transf.	13.45	16.76	8.52	8.59

Table : A10

corresponding to Section: 6.3

Experiment : 2

Sentence : de man koopt een tas

Choice model: Restle

Restrictions: none except positivity of weights

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.188	.229	.171	.160
$noun_1$	1.000	1.000	1.000	1.000
verb	1.228	2.163	2.831	.352
art_2	.195	.125	.340	.150
$noun_2$	.572	.449	1.004	.554
MAD-values				
overall	.135	.155	.102	.103
det. PWCs	.152	.187	.087	.114
prob. PWCs	.116	.118	.110	.090
L-values	125.09	134.95	160.28	85.64
df	17	17	25	17
p				
z-transf.	10.07	10.68	10.90	7.34

Table : A11

corresponding to Section: 6.3

Experiment : 2

Sentence : de man koopt een tas

Choice model: Luce

Restrictions: $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.179	.133	.000	.109
$noun_1$	1.000	1.000	1.000	1.000
verb	2.275	2.125	3.797	.454
art_2				
$noun_2$				
MAD-values				
overall	.128	.144	.124	.084
det. PWCs	.087	.087	.087	.087
prob. PWCs	.150	.176	.144	.082
L-values	292.80	355.08	229.65	93.80
df	27	27	27	27
p				
z.transf.	16.92	19.37	14.15	6.41

Table : A12
corresponding to Section: 6.3
Experiment : 2
Sentence : de man koopt een tas
Choice model: Thurstone
Restrictions: $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.002	.000	.048	.642
$noun_1$	.645	.555	.692	1.292
verb	.985	.966	1.248	.000
art_2				
$noun_2$				
MAD-values				
overall	.106	.132	.081	.084
det. PWCs	-	-	-	-
prob. PWCs	.106	.132	.081	.084
L-values	263.46	368.51	169.09	175.37
df	44	44	44	44
p				
z-transf.	13.63	17.82	9.06	9.40

Table : A13
corresponding to Section: 6.3
Experiment : 2
Sentence : de man koopt een tas
Choice model: Restle
Restrictions: $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$

Syntactic options	CGV	CGN	CGR	CGP
Weights				
art_1	.192	.180	.178	.147
$noun_1$	1.000	1.000	1.000	1.000
verb	2.315	2.701	3.429	.576
art_2				
$noun_2$				
MAD-values				
overall	.135	.155	.102	.103
det. PWCs	.152	.187	.087	.114
prob. PWCs	.116	.118	.110	.090
L-values	147.81	151.22	193.85	113.84
df	19	19	27	19
p				
z.transf.	11.11	11.31	12.41	9.01

Table : A14 Residual analyses
corresponding to Section: 6.3 (Page 163)
Experiment : 2
Sentence : de man koopt een tas

Choice mod:	Thurstone		Luce	Restle
Restrict.:	/verb/ $\succ$ /art/	/verb/ $\succ$ /noun/	/verb/ $\succ$ /noun/	/verb/ $\succ$ /noun/
Syntactic options	CGP	CGP	CGP	CGP
Weights				
art_1	.362	.142	.125	.147
$noun_1$	1.048	.733	1.000	1.000
verb	.362	.733	1.000	1.000
art_2				
$noun_2$				
MAD-values				
overall	.097	.117	.089	.107
det. PWCs	-	-	.090	.114
prob. PWCs	.097	.117	.087	.099
L-values	212.98	289.91	110.01	120.74
df	44	44	27	19
p				
z-transf.	11.31	14.75	7.55	9.46

Table : A15 Residual analyses
corresponding to Section: 6.3 (Pages 171 and 173)
Experiment : 2
Sentence : de man koopt een tas

Choice mod:	Luce	Luce
Restrict. :	/noun/ $\prec$ /verb/ ; κ	/noun/ $\prec$ /verb/; κ,sem
Syntactic options	CGP	CGP
Weights		
art_1	.104	.111
$noun_1$	1.000	1.000
verb	1.000	1.696
κ	.055	.053
sem	-	- 1.307
MAD-values		
overall	.067	.053
det. PWCs	-	-
prob. PWCs	.067	.053
L-values	119.87	75.931
df	42	41
p		$.967 \times 10^{-3}$
z.transf.	6.37	3.32

Table A16 **Composition of L and MAD for CGP-Luce under restrictions $/art_1/ = /art_2/ < /noun_1/ = /noun_2/ \leqslant /verb/$, applied to the data of Experiment 2 (Page 163)**

PWC NR:	PAIR-NUMBERS	WORDPAIRS (X,Y) - (W,Z)	EXPECTED FREQ. >	EXPECTED FREQ. <	OBSERVED FREQ. >	OBSERVED FREQ. <	CONTRIB. TO L	ABS.DEV. PDAT-PMOD
1	1- 2	D,M - D,K	49.00	0.0	47.00	2.00		0.0408
2	1- 3	D,M - D,E	49.00	0.0	48.00	1.00		0.0204
3	1- 4	D,M - D,T	49.00	0.0	46.00	3.00		0.0612
4	1- 5	D,M - M,K	49.00	0.0	37.50	11.50		0.2347
5	1- 6	D,M - M,E	49.00	0.0	46.50	2.50		0.0510
6	1- 7	D,M - M,T	49.00	0.0	46.50	2.50		0.0510
7	1- 8	D,M - K,E	49.00	0.0	45.00	4.00		0.0816
8	1- 9	D,M - K,T	49.00	0.0	39.50	9.50		0.1939
9	1-10	D,M - E,T	24.50	24.50	27.50	21.50	0.7365	0.0612
10	2- 3	D,K - D,E	36.75	12.25	32.50	16.50	1.8402	0.0867
11	2- 4	D,K - D,T	33.32	15.68	35.50	13.50	0.4580	0.0445
12	2- 5	D,K - M,K	5.44	43.56	2.00	47.00	3.1425	0.0702
13	2- 6	D,K - M,E	33.32	15.68	31.00	18.00	0.4927	0.0473
14	2- 7	D,K - M,T	27.22	21.78	15.00	34.00	12.4082	0.2494
15	2- 8	D,K - K,E	24.50	24.50	13.50	35.50	10.2395	0.2245
16	2- 9	D,K - K,T	5.44	43.56	1.50	47.50	4.3614	0.0804
17	2-10	D,K - E,T	0.0	49.00	3.50	45.50		0.0714
18	3- 4	D,E - D,T	20.32	28.68	26.50	22.50	3.1564	0.1262
19	3- 5	D,E - M,K	1.96	47.04	2.00	47.00	0.0009	0.0009
20	3- 6	D,E - M,E	20.32	28.68	20.50	28.50	0.0028	0.0037
21	3- 7	D,E - M,T	14.41	34.59	9.50	39.50	2.5701	0.1002
22	3- 8	D,E - K,E	12.25	36.75	6.50	42.50	4.1177	0.1173
23	3- 9	D,E - K,T	1.96	47.04	4.00	45.00	1.7202	0.0417
24	3-10	D,E - E,T	0.0	49.00	1.50	47.50		0.0306
25	4- 5	D,T - M,K	2.72	46.28	4.00	45.00	0.5610	0.0261
26	4- 6	D,T - M,E	24.50	24.50	21.00	28.00	1.0034	0.0714
27	4- 7	D,T - M,T	18.15	30.85	7.00	42.00	12.5717	0.2275
28	4- 8	D,T - K,E	15.68	33.32	8.50	40.50	5.3980	0.1465
29	4- 9	D,T - K,T	2.72	46.28	4.50	44.50	1.0403	0.0363
30	4-10	D,T - E,T	0.0	49.00	2.00	47.00		0.0408
31	5- 6	M,K - M,E	46.28	2.72	45.50	3.50	0.2181	0.0159
32	5- 7	M,K - M,T	44.55	4.45	43.00	6.00	0.5395	0.0316
33	5- 8	M,K - K,E	43.56	5.44	42.00	7.00	0.4663	0.0318
34	5- 9	M,K - K,T	24.50	24.50	27.00	22.00	0.5111	0.0510
35	5-10	M,K - E,T	0.0	49.00	11.00	38.00		0.2245
36	6- 7	M,E - M,T	18.15	30.85	4.00	45.00	21.8702	0.2887
37	6- 8	M,E - K,E	15.68	33.32	5.00	44.00	13.0384	0.2180
38	6- 9	M,E - K,T	2.72	46.28	4.00	45.00	0.5610	0.0261
39	6-10	M,E - E,T	0.0	49.00	2.50	46.50		0.0510
40	7- 8	M,T - K,E	21.78	27.22	25.50	23.50	1.1353	0.0759
41	7- 9	M,T - K,T	4.45	44.55	10.00	39.00	5.8111	0.1132
42	7-10	M,T - E,T	0.0	49.00	3.00	46.00		0.0612
43	8- 9	K,E - K,T	5.44	43.56	5.00	44.00	0.0410	0.0090
44	8-10	K,E - E,T	0.0	49.00	2.00	47.00		0.0408
45	9-10	K,T - E,T	0.0	49.00	7.00	42.00		0.1429

L	MAD
110.0135	0.0894

Table A17 Composition of L and MAD for CGR-Thurstone under restrictions $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$, applied to the data of Experiment 2 (Page 163)

PWC NR:	PAIR-NUMBERS	WORDPAIRS (X,Y)- (W,Z)	EXPECTED FREQ. >	EXPECTED FREQ. <	OBSERVED FREQ. >	OBSERVED FREQ. <	CONTRIB. TO L	ABS.DEV. PDAT-PMOD
1	1- 2	D,M - D,K	45.28	3.72	47.00	2.00	1.0206	0.0351
2	1- 3	D,M - D,E	48.78	0.22	48.00	1.00	1.5046	0.0160
3	1- 4	D,M - D,T	47.86	1.14	46.00	3.00	2.1711	0.0380
4	1- 5	D,M - M,K	38.46	10.54	37.50	11.50	0.1092	0.0196
5	1- 6	D,M - M,E	47.86	1.14	46.50	2.50	1.2544	0.0278
6	1- 7	D,M - M,T	44.61	4.39	46.50	2.50	1.0399	0.0385
7	1- 8	D,M - K,E	45.28	3.72	45.00	4.00	0.0226	0.0057
8	1- 9	D,M - K,T	38.46	10.54	39.50	9.50	0.1339	0.0212
9	1-10	D,M - E,T	24.50	24.50	27.50	21.50	0.7365	0.0612
10	2- 3	D,K - D,E	43.35	5.65	32.50	16.50	16.6276	0.2214
11	2- 4	D,K - D,T	34.83	14.17	35.50	13.50	0.0448	0.0136
12	2- 5	D,K - M,K	12.72	36.28	2.00	47.00	16.9346	0.2188
13	2- 6	D,K - M,E	34.83	14.17	31.00	18.00	1.3913	0.0782
14	2- 7	D,K - M,T	22.77	26.23	15.00	34.00	5.1271	0.1587
15	2- 8	D,K - K,E	24.50	24.50	13.50	35.50	10.2395	0.2245
16	2- 9	D,K - K,T	12.72	36.28	1.50	47.50	19.1857	0.2290
17	2-10	D,K - E,T	3.72	45.28	3.50	45.50	0.0141	0.0045
18	3- 4	D,E - D,T	12.72	36.28	26.50	22.50	17.4017	0.2812
19	3- 5	D,E - M,K	1.59	47.41	2.00	47.00	0.1028	0.0084
20	3- 6	D,E - M,E	12.72	36.28	20.50	28.50	5.8097	0.1588
21	3- 7	D,E - M,T	4.85	44.15	9.50	39.50	3.9869	0.0949
22	3- 8	D,E - K,E	5.65	43.35	6.50	42.50	0.1376	0.0173
23	3- 9	D,E - K,T	1.59	47.41	4.00	45.00	2.6944	0.0492
24	3-10	D,E - E,T	0.22	48.78	1.50	47.50	3.2716	0.0262
25	4- 5	D,T - M,K	5.65	43.35	4.00	45.00	0.6013	0.0337
26	4- 6	D,T - M,E	24.50	24.50	21.00	28.00	1.0034	0.0714
27	4- 7	D,T - M,T	12.72	36.28	7.00	42.00	3.9361	0.1167
28	4- 8	D,T - K,E	14.17	34.83	8.50	40.50	3.5271	0.1157
29	4- 9	D,T - K,T	5.65	43.35	4.50	44.50	0.2836	0.0235
30	4-10	D,T - E,T	1.14	47.86	2.00	47.00	0.5498	0.0176
31	5- 6	M,K - M,E	43.35	5.65	45.50	3.50	1.0555	0.0439
32	5- 7	M,K - M,T	34.83	14.17	43.00	6.00	7.8074	0.1667
33	5- 8	M,K - K,E	36.28	12.72	42.00	7.00	3.9361	0.1167
34	5- 9	M,K - K,T	24.50	24.50	27.00	22.00	0.5111	0.0510
35	5-10	M,K - E,T	10.54	38.46	11.00	38.00	0.0254	0.0094
36	6- 7	M,E - M,T	12.72	36.28	4.00	45.00	10.1305	0.1780
37	6- 8	M,E - K,E	14.17	34.83	5.00	44.00	10.1472	0.1871
38	6- 9	M,E - K,T	5.65	43.35	4.00	45.00	0.6013	0.0337
39	6-10	M,E - E,T	1.14	47.86	2.50	46.50	1.2544	0.0278
40	7- 8	M,T - K,E	26.23	22.77	25.50	23.50	0.0431	0.0148
41	7- 9	M,T - K,T	14.17	34.83	10.00	39.00	1.8485	0.0851
42	7-10	M,T - E,T	4.39	44.61	3.00	46.00	0.5357	0.0283
43	8- 9	K,E - K,T	12.72	36.28	5.00	44.00	7.6397	0.1575
44	8-10	K,E - E,T	3.72	45.28	2.00	47.00	1.0206	0.0351
45	9-10	K,T - E,T	10.54	38.46	7.00	42.00	1.6661	0.0722

L	MAD
169.0860	0.0808

Table A18 Sources of deficiencies of CGP-Luce (under restriction $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/ \leqslant /verb/$) applied to the *declarative* version's data; specified for *word pairs*

	(a) sum of contributions to L		(b) MAD	(c) average overprediction
de man	.74	n = 1	.088	+ .075
de koopt	32.94	7	.102	+ .067
de een	13.41	7	.059	- .010
de tas	24.19	7	.087	+ .050
man koopt	5.44	7	.067	- .053
man een	37.18	7	.086	+ .028
man tas	56.89	7	.133	- .133
koopt een	34.43	7	.105	- .086
koopt tas	14.04	7	.077	- .018
een tas	.74	1	.080	+ .080

Table A19 Sources of deficiencies of CGR-Thurstone (under restriction $/art_1/ = /art_2/$ and $/noun_1 = /noun_2$) applied to the *declarative* version's data; specified for *word pairs*

	(a) sum of contributions to L		(b) MAD	(c) average overprediction
de man	7.99	n = 9	.029	- .005
de koopt	70.58	9	.132	+ .128
de een	51.54	9	.097	- .097
de tas	29.52	9	.079	+ .067
man koopt	31.08	9	.079	- .072
man een	32.65	9	.090	+ .044
man tas	34.45	9	.098	- .003
koopt een	36.71	9	.097	- .025
koopt tas	34.56	9	.080	- .037
een tas	9.07	9	.031	.000

Table : B1
corresponding to Section: 7.1
Experiment : 3
Sentence : koopt de man een tas ?
Choice model: Luce
Restrictions: none except positivity weights; κ

Syntactic options	CGV	CGN	CGR	CGP
Weights : κ	.032	.108	.182	.019
art_1	.000	.188	.000	.094
$noun_1$	1.000	1.000	1.000	1.000
verb	.505	1.559	4.178	.150
art_2	.048	.000	.000	.049
$noun_2$	.340	1.070	1.977	.584
MAD-values				
overall	.090	.139	.112	.046
det. PWCs	-	-	-	-
prob. PWCs	.090	.139	.112	.046
L-values	190.27	414.44	265.43	59.41
df	40	40	40	40
p				$.244 \times 10^{-1}$
z-transf.	10.62	19.90	14.15	2.01

Table : B2
corresponding to Section: 7.1
Experiment : 3
Sentence : (dat) de man een tas koopt
Choice model: Luce
Restrictions: none except positivity weights; κ

Syntactic options	CGV	CGN	CGR	CGP
Weights : κ	.134	.108	.098	.052
art_1	.000	.061	.000	.053
$noun_1$	1.000	1.000	1.000	1.000
verb	.818	.849	1.299	.074
art_2	.171	.000	.000	.267
$noun_2$	.553	1.385	1.995	1.175
MAD-values				
overall	.136	.086	.091	.053
det. PWCs	-	-	-	-
prob. PWCs	.136	.086	.091	.053
L-values	404.66	156.22	197.99	71.59
df	40	40	40	40
p				$.183 \times 10^{-2}$
z.transf.	19.56	8.79	11.01	3.08

Table : B3

corresponding to Section: 7.1

Experiment : 3

Sentence : koopt de man een tas ?

Choice model: Luce

Restrictions: $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$; κ

Syntactic options	CGV	CGN	CGR	CGP
Weights: κ	.078	.130	.125	.026
art_1	.079	.105	.000	.091
$noun_1$	1.000	1.000	1.000	1.000
verb	1.264	1.530	2.769	.193
$art_2 = art_1$				
$noun_2 = noun_1$				
MAD-values				
overall	.102	.144	.112	.054
det. PWCs	-	-	-	-
prob. PWCs	.102	.144	.122	.054
L-values	246.84	445.82	274.41	83.73
df	42	42	42	42
p				$.259 \times 10^{-3}$
z-transf.	13.11	20.75	14.32	3.83

Table : B4

corresponding to Section: 7.1

Experiment : 3

Sentence : (dat) de man een tas koopt

Choice model: Luce

Restrictions: $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$; κ

Syntactic options	CGV	CGN	CGR	CGP
Weights: κ	.286	.109	.186	.054
art_1	.108	.031	.000	.134
$noun_1$	1.000	1.000	1.000	1.000
verb	1.387	.843	2.088	.104
$art_2 = art_1$				
$noun_2 = noun_1$				
MAD-values				
overall	.140	.087	.103	.065
det. PWCs	-	-	-	-
prob. PWCs	.140	.087	.103	.065
L-values	453.66	166.22	250.49	118.47
df	42	42	42	42
p				
z.transf.	21.01	9.12	13.27	6.28

Table : B5 (Residual analyses) corresponding to Section : 7.1 (Page 186) Experiment : 3			
Sentence :	interrogative	object clause	
Choice model:	Luce	Luce	
Restrictions:	/verb/ ≥ /noun/ ; κ	/verb/ ≥ /noun/ ; κ	/verb/ ≥ /noun/ ; κ
Syntactic options	CGP	CGN	CGP
Weights : κ	.040	.117	.091
art_1	.118	.030	.159
$noun_1$	1.000	1.000	1.000
verb	1.000	1.000	1.000
$art_2 = art_1$			
$noun_2 = noun_1$			
MAD-values			
overall	.074	.087	.083
deterministic PWCs	-	-	-
probabilistic PWCs	.074	.087	.083
L-values	144.95	167.48	196.34
df	42	42	42
p			
z-transformation	7.92	9.19	10.70

Table B6 Composition of L and MAD for CGP-Luce under restrictions $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/$, with additional parameters κ and *sem*, applied to the *interrogative* version's data (Page 188)

PWC NR:	PAIR-NUMBERS	WORDPAIRS (X,Y)- (W,Z)	EXPECTED FREQ. >	EXPECTED FREQ. <	OBSERVED FREQ. >	OBSERVED FREQ. <	CONTRIB. TO L	ABS.DEV. PDAT-PMOD
1	1- 2	D,M - D,K	48.54	1.46	49.00	1.00	0.1684	0.0092
2	1- 3	D,M - D,E	49.50	0.50	48.00	2.00	2.5763	0.0300
3	1- 4	D,M - D,T	49.30	0.70	49.00	1.00	0.1113	0.0059
4	1- 5	D,M - M,K	41.18	8.82	45.00	5.00	2.3091	0.0764
5	1- 6	D,M - M,E	49.30	0.70	47.00	3.00	4.2105	0.0459
6	1- 7	D,M - M,T	46.48	3.52	46.00	4.00	0.0666	0.0095
7	1- 8	D,M - K,E	48.54	1.46	49.00	1.00	0.1684	0.0092
8	1- 9	D,M - K,T	41.18	8.82	40.50	9.50	0.0622	0.0136
9	1-10	D,M - E,T	25.00	25.00	27.00	23.00	0.3203	0.0400
10	2- 3	D,K - D,E	37.39	12.61	30.00	20.00	5.2414	0.1479
11	2- 4	D,K - D,T	33.90	16.10	35.00	15.00	0.1113	0.0219
12	2- 5	D,K - M,K	6.16	43.84	4.00	46.00	0.9716	0.0432
13	2- 6	D,K - M,E	33.90	16.10	29.50	20.50	1.7071	0.0881
14	2- 7	D,K - M,T	14.21	35.79	12.00	38.00	0.4975	0.0442
15	2- 8	D,K - K,E	25.00	25.00	14.00	36.00	10.0194	0.2200
16	2- 9	D,K - K,T	6.16	43.84	3.00	47.00	2.2273	0.0632
17	2-10	D,K - E,T	1.46	48.54	1.00	49.00	0.1684	0.0092
18	3- 4	D,E - D,T	20.76	29.24	24.50	25.50	1.1343	0.0747
19	3- 5	D,E - M,K	2.26	47.74	4.50	45.50	1.8207	0.0448
20	3- 6	D,E - M,E	20.76	29.24	21.50	28.50	0.0445	0.0147
21	3- 7	D,E - M,T	5.90	44.10	7.50	42.50	0.4556	0.0319
22	3- 8	D,E - K,E	12.61	37.39	7.00	43.00	3.7793	0.1121
23	3- 9	D,E - K,T	2.26	47.74	2.50	47.50	0.0254	0.0048
24	3-10	D,E - E,T	0.50	49.50	0.0	50.00	1.0100	0.0100
25	4- 5	D,T - M,K	3.13	46.87	6.00	44.00	2.2532	0.0575
26	4- 6	D,T - M,E	25.00	25.00	23.50	26.50	0.1801	0.0300
27	4- 7	D,T - M,T	7.93	42.07	5.50	44.50	0.9733	0.0486
28	4- 8	D,T - K,E	16.10	33.90	11.50	38.50	2.0550	0.0919
29	4- 9	D,T - K,T	3.13	46.87	5.00	45.00	1.0230	0.0375
30	4-10	D,T - E,T	0.70	49.30	0.50	49.50	0.0670	0.0041
31	5- 6	M,K - M,E	46.87	3.13	46.00	4.00	0.2399	0.0175
32	5- 7	M,K - M,T	36.93	13.07	38.50	11.50	0.2626	0.0314
33	5- 8	M,K - K,E	43.84	6.16	40.00	10.00	2.3548	0.0768
34	5- 9	M,K - K,T	25.00	25.00	21.00	29.00	1.2855	0.0800
35	5-10	M,K - E,T	8.82	41.18	7.00	43.00	0.4842	0.0364
36	6- 7	M,E - M,T	7.93	42.07	6.00	44.00	0.6009	0.0386
37	6- 8	M,E - K,E	16.10	33.90	11.00	39.00	2.5468	0.1019
38	6- 9	M,E - K,T	3.13	46.87	2.00	48.00	0.4935	0.0225
39	6-10	M,E - E,T	0.70	49.30	1.00	49.00	0.1113	0.0059
40	7- 8	M,T - K,E	35.79	14.21	34.50	15.50	0.1604	0.0258
41	7- 9	M,T - K,T	13.07	36.93	9.00	41.00	1.8577	0.0814
42	7-10	M,T - E,T	3.52	46.48	5.00	45.00	0.5941	0.0295
43	8- 9	K,E - K,T	6.16	43.84	2.00	48.00	4.2053	0.0832
44	8-10	K,E - E,T	1.46	48.54	1.00	49.00	0.1684	0.0092
45	9-10	K,T - E,T	8.82	41.18	10.00	40.00	0.1851	0.0236

L	MAD
61.3086	0.0468

Table B7 **Composition of L and MAD for CGN-Luce under restrictions /art$_1$/ = /art$_2$/ ; /noun$_1$/ = /noun$_2$/ $\leqslant$ /verb/, with additional parameters κ and *sem*, applied to the *object clause* version's data (Page 188)**

PWC NR:	PAIR-NUMBERS	WORDPAIRS (X,Y)- (W,Z)	EXPECTED FREQ. >	EXPECTED FREQ. <	OBSERVED FREQ. >	OBSERVED FREQ. <	CONTRIB. TO L	ABS.DEV. PDAT-PMOD
1	1- 2	D,M - D,K	44.29	4.71	48.00	1.00	4.6287	0.0758
2	1- 3	D,M - D,E	47.66	1.34	48.00	1.00	0.0973	0.0069
3	1- 4	D,M - D,T	47.29	1.71	45.00	4.00	2.3238	0.0467
4	1- 5	D,M - M,K	29.01	19.99	38.50	10.50	8.2649	0.1936
5	1- 6	D,M - M,E	47.29	1.71	48.50	0.50	1.2250	0.0247
6	1- 7	D,M - M,T	42.74	6.26	41.50	7.50	0.2682	0.0254
7	1- 8	D,M - K,E	46.45	2.55	46.00	3.00	0.0809	0.0093
8	1- 9	D,M - K,T	44.66	4.34	40.00	9.00	4.3209	0.0952
9	1-10	D,M - E,T	24.50	24.50	25.50	23.50	0.0817	0.0204
10	2- 3	D,K - D,E	38.76	10.24	32.00	17.00	4.9648	0.1379
11	2- 4	D,K - D,T	36.56	12.44	33.00	16.00	1.2908	0.0726
12	2- 5	D,K - M,K	6.56	42.44	0.0	49.00	14.0793	0.1338
13	2- 6	D,K - M,E	36.56	12.44	31.00	18.00	3.0706	0.1134
14	2- 7	D,K - M,T	20.63	28.37	11.00	38.00	8.3694	0.1964
15	2- 8	D,K - K,E	32.34	16.66	26.50	22.50	2.9701	0.1192
16	2- 9	D,K - K,T	25.62	23.38	15.50	33.50	8.5253	0.2066
17	2-10	D,K - E,T	4.71	44.29	1.50	47.50	3.2194	0.0656
18	3- 4	D,E - D,T	21.42	27.58	22.50	26.50	0.0967	0.0221
19	3- 5	D,E - M,K	1.92	47.08	2.50	46.50	0.1656	0.0118
20	3- 6	D,E - M,E	21.42	27.58	22.50	26.50	0.0967	0.0221
21	3- 7	D,E - M,T	7.90	41.10	9.50	39.50	0.3691	0.0327
22	3- 8	D,E - K,E	16.62	32.38	15.00	34.00	0.2421	0.0330
23	3- 9	D,E - K,T	11.01	37.99	5.50	43.50	4.1436	0.1124
24	3-10	D,E - E,T	1.34	47.66	0.0	49.00	2.7182	0.0274
25	4- 5	D,T - M,K	2.45	46.55	8.00	41.00	8.5345	0.1133
26	4- 6	D,T - M,E	24.50	24.50	25.00	24.00	0.0204	0.0102
27	4- 7	D,T - M,T	9.72	39.28	11.00	38.00	0.2046	0.0262
28	4- 8	D,T - K,E	19.50	29.50	16.50	32.50	0.7803	0.0611
29	4- 9	D,T - K,T	13.31	35.69	10.00	39.00	1.2012	0.0676
30	4-10	D,T - E,T	1.71	47.29	3.50	45.50	1.4967	0.0365
31	5- 6	M,K - M,E	46.55	2.45	46.50	2.50	0.0012	0.0011
32	5- 7	M,K - M,T	40.41	8.59	41.00	8.00	0.0499	0.0120
33	5- 8	M,K - K,E	45.39	3.61	47.50	1.50	1.6839	0.0431
34	5- 9	M,K - K,T	42.95	6.05	42.00	7.00	0.1619	0.0193
35	5-10	M,K - E,T	19.99	29.01	12.50	36.50	5.0233	0.1528
36	6- 7	M,E - M,T	9.72	39.28	9.00	40.00	0.0674	0.0146
37	6- 8	M,E - K,E	19.50	29.50	14.50	34.50	2.2079	0.1020
38	6- 9	M,E - K,T	13.31	35.69	7.00	42.00	4.6822	0.1288
39	6-10	M,E - E,T	1.71	47.29	0.50	48.50	1.2250	0.0247
40	7- 8	M,T - K,E	35.65	13.35	32.50	16.50	0.9794	0.0643
41	7- 9	M,T - K,T	29.46	19.54	24.00	25.00	2.4837	0.1115
42	7-10	M,T - E,T	6.26	42.74	7.00	42.00	0.0977	0.0151
43	8- 9	K,E - K,T	17.68	31.32	6.00	43.00	14.2911	0.2384
44	8-10	K,E - E,T	2.55	46.45	6.50	42.50	4.6225	0.0807
45	9-10	K,T - E,T	4.34	44.66	6.50	42.50	1.0412	0.0442

L	MAD
126.4690	0.0705

Table B8 Composition of L and MAD for CGP-Luce under restrictions $/art_1/ = /art_2/$; $/noun_1/ = /noun_2/ \leqslant /verb/$, with additional parameters κ and *sem*, applied to the *object clause* version's data (Page 188)

PWC NR:	PAIR-NUMBERS	WORDPAIRS (X,Y)- (W,Z)	EXPECTED FREQ. >	EXPECTED FREQ. <	OBSERVED FREQ. >	OBSERVED FREQ. <	CONTRIB. TO L	ABS.DEV. PDAT-PMOD
1	1- 2	D,M - D,K	46.03	2.97	48.00	1.00	1.8523	0.0403
2	1- 3	D,M - D,E	47.92	1.08	48.00	1.00	0.0060	0.0016
3	1- 4	D,M - D,T	47.53	1.47	45.00	4.00	3.1004	0.0517
4	1- 5	D,M - M,K	38.07	10.93	38.50	10.50	0.0223	0.0088
5	1- 6	D,M - M,E	47.53	1.47	48.50	0.50	0.8755	0.0197
6	1- 7	D,M - M,T	42.93	6.07	41.50	7.50	0.3596	0.0291
7	1- 8	D,M - K,E	46.03	2.97	46.00	3.00	0.0002	0.0005
8	1- 9	D,M - K,T	38.07	10.93	40.00	9.00	0.4603	0.0394
9	1-10	D,M - E,T	24.50	24.50	25.50	23.50	0.0817	0.0204
10	2- 3	D,K - D,E	36.34	12.66	32.00	17.00	1.8832	0.0886
11	2- 4	D,K - D,T	33.17	15.83	33.00	16.00	0.0028	0.0035
12	2- 5	D,K - M,K	9.00	40.00	0.0	49.00	19.8902	0.1837
13	2- 6	D,K - M,E	33.17	15.83	31.00	18.00	0.4307	0.0443
14	2- 7	D,K - M,T	15.36	33.64	11.00	38.00	1.9205	0.0891
15	2- 8	D,K - K,E	24.50	24.50	26.50	22.50	0.3269	0.0408
16	2- 9	D,K - K,T	9.00	40.00	15.50	33.50	4.9693	0.1326
17	2-10	D,K - E,T	2.97	46.03	1.50	47.50	0.9418	0.0301
18	3- 4	D,E - D,T	20.68	28.32	22.50	26.50	0.2761	0.0372
19	3- 5	D,E - M,K	3.56	45.44	2.50	46.50	0.3783	0.0217
20	3- 6	D,E - M,E	20.68	28.32	22.50	26.50	0.2761	0.0372
21	3- 7	D,E - M,T	6.73	42.27	9.50	39.50	1.1986	0.0566
22	3- 8	D,E - K,E	12.66	36.34	15.00	34.00	0.5626	0.0478
23	3- 9	D,E - K,T	3.56	45.44	5.50	43.50	0.9869	0.0396
24	3-10	D,E - E,T	1.08	47.92	0.0	49.00	2.1817	0.0220
25	4- 5	D,T - M,K	4.75	44.25	8.00	41.00	2.0839	0.0663
26	4- 6	D,T - M,E	24.50	24.50	25.00	24.00	0.0204	0.0102
27	4- 7	D,T - M,T	8.77	40.23	11.00	38.00	0.6511	0.0455
28	4- 8	D,T - K,E	15.83	33.17	16.50	32.50	0.0418	0.0137
29	4- 9	D,T - K,T	4.75	44.25	10.00	39.00	5.0353	0.1071
30	4-10	D,T - E,T	1.47	47.53	3.50	45.50	2.1125	0.0415
31	5- 6	M,K - M,E	44.25	4.75	46.50	2.50	1.4043	0.0459
32	5- 7	M,K - M,T	32.83	16.17	41.00	8.00	6.9663	0.1668
33	5- 8	M,K - K,E	40.00	9.00	47.50	1.50	10.9521	0.1531
34	5- 9	M,K - K,T	24.50	24.50	42.00	7.00	27.7370	0.3571
35	5-10	M,K - E,T	10.93	38.07	12.50	36.50	0.2801	0.0320
36	6- 7	M,E - M,T	8.77	40.23	9.00	40.00	0.0074	0.0047
37	6- 8	M,E - K,E	15.83	33.17	14.50	34.50	0.1672	0.0271
38	6- 9	M,E - K,T	4.75	44.25	7.00	42.00	1.0440	0.0459
39	6-10	M,E - E,T	1.47	47.53	0.50	48.50	0.8755	0.0197
40	7- 8	M,T - K,E	33.64	15.36	32.50	16.50	0.1207	0.0232
41	7- 9	M,T - K,T	16.17	32.83	24.00	25.00	5.3288	0.1598
42	7-10	M,T - E,T	6.07	42.93	7.00	42.00	0.1549	0.0189
43	8- 9	K,E - K,T	9.00	40.00	6.00	43.00	1.3547	0.0612
44	8-10	K,E - E,T	2.97	46.03	6.50	42.50	3.3888	0.0720
45	9-10	K,T - E,T	10.93	38.07	6.50	42.50	2.6032	0.0905
							L 115.3142	MAD 0.0589

INDEX OF SYMBOLS, ABBREVIATIONS AND NOTIONS

This index lists the most frequently used symbols, abbreviations and notions together with the place of their introduction. Symbols etc. which have only local relevance have been omitted as their meaning will be apparent from the immediate context of their occurrence.

Note: The words constituting the various model sentences in Chapters 2 to 7 are often reduced to their initials, in capitals. The corresponding lower cases, from Chapter 5 onward, are used in order to refer to the weights of these words.

REFERENCES

Anderson, J.M., *The grammar of case. Towards a localistic theory.* Cambridge: Cambridge University Press, 1971.

Barber, T.X., Pitfalls in research: nine investigator and experimenter effects. In: *Second handbook of research on teaching*, R.M.W. Travers, ed. Chicago: Rand McNally, 1973.

Beals, R., Krantz, D.H. & Tversky, A., Foundations of multidimensional scaling. *Psychological Review, 75*, 1968, 127-142.

Bezembinder, T.G.G., *Van rangorde naar continuum*. Deventer: Van Loghum-Slaterus, 1970.

Blalock, H.M., Jr., The measurement problem: a gap between the languages of theory and research. In: *Methodology in social research*, Hubert M. Blalock Jr. & Ann B. Blalock (eds.). London: McGraw-Hill, 1971.

Bresnan, J., A realistic transformational grammar. In: *Linguistic theory and psychological reality*, M. Halle, J. Bresnan & G.A. Miller, eds. Cambridge, Mass.: The MIT Press, 1968.

Carden, C., Disambiguation, favored readings and variable rules. In: *New ways of analyzing variation in English*, C.J. Bailey & R.W. Shuy, eds. Washington: Georgetown University Press, 1973.

Chomsky, N., Formal properties of grammars. In: *Handbook of mathematical psychology, Vol. 2*, R.D. Luce, R.R. Bush & E. Galanter, eds. New York: Wiley, 1963.

Chomsky, N., *Aspects of the theory of syntax*. Cambridge, Mass.: The MIT Press, 1965.

Coombs, C.H., Theory and methods of social measurement. In: *Research methods in the behavioral sciences*, L. Festinger & D. Katz, eds. New York: Rinehart and Winston, 1953.

Coombs, C.H., *A theory of data*. New York: Wiley, 1964.

Elliot, D., Legum,S. & Thompson, S., Syntactic variation as linguistic data, *Papers from the fifth regional meeting of the Chicago Linguistic Society*, 1969.

Fillmore, C.J., The case for case. In: *Universals in linguistic theory*, E. Bach & R.T. Harms, eds. New York: Holt, Rinehart and Winston, 1968.

Fitialov, S.J., On the equivalence of IC grammars and dependency grammars. In: *Mathematical models of language, vol. III*, Ferenc Kiefer, ed. Frankfurt am Main: Athenaeum Verlag, 1973.

Fodor, J.A., Garret, M.F., Walker, E.C.T. & Parkes, C.H., Against definitions. *Cognition, 8*, 1980, 263-367.

Fodor, J.A., Bever, T. & Garret, M., *The psychology of Language: an introduction to psycholinguistics and generative grammar*. New York. McGraw-Hill, 1974.

Gaifman, C., Dependency systems and phrase structure systems. *Information and control, 8*, 1965, 304-337.

Greenbaum, S., Informant elicitation of data on syntactic variation. *Lingua, 31*, 1973, 201-212.

Greenbaum, S., *Acceptability in Language*. Den Haag: Mouton, 1977.

Grinder, J. & Postal, P., Missing antecedents. *Linguistic Inquiry, 2*, 1971, 209-312.

Halliday, M.A.K. & Hasan, R., *Cohesion in English*. London. Longmans, 1977.

Harary, F., Norman, R.Z. & Cartwright, D., *Structural models: an introduction to the theory of directed graphs*. New York: Wiley, 1965.

Hays, D.G., *Introduction to computational linguistics*. New York: Elsevier, 1967.

Hays, D.G., Dependency theory: a formalism and some observations. *Language, 33*, 1964, 283-340.

James, F. & Roos, M., MINUIT, a system for function minimization and analysis of the parameter errors and correlations. *Computer physics communications, 10*, 1975, 343-367.

Johnson, S.C., Hierarchical clustering schemes. *Psychometrika, 32*, 1967, 241-254.

Kaplan, A., *The conduct of inquiry; methodology for behavioral science*. New York: Chandler, 1964.

Kaplan, R., A general syntactic processor. In: *Natural language processing*, R. Rustin, ed Englewoods Cliffs, N.J. Prentice Hall, 1973.

Labov, W., *What is a linguistic fact ?* Lisse: The Peter de Ridder Press, 1975.

Legum, S., Elliot, D.E. & Thompson, S., Considerations in the analysis of syntactic variation. *SWRL Professional Paper* , Los Alamitos, 1974.

Levelt, W.J.M., The scaling of syntactic relatedness. A new method in psycholinguistic research. *Psychonomic Science, 17*, 1969, 351-352.

Levelt, W.J.M., A scaling approach to the study of syntactic relations. In: *Advances in psycholinguistics*, G.B. Flores d'Arcais & W.J.M. Levelt, eds., Amsterdam: North Holland, 1970.

Levelt, W.J.M., *Formal grammars in linguistics and psycholinguistics, Vol. 1: An introduction to the theory of formal languages and automata.* Den Haag: Mouton, 1974a.

Levelt, W.J.M., *Formal grammars in linguistics and psycholinguistics, Vol. 2: Applications in linguistic theory.* Den Haag: Mouton, 1974b.

Levelt, W.J.M., *Formal grammars in linguistics and psycholinguistics, Vol. 3: Psycholinguistic applications.* Den Haag: Mouton, 1974c.

Loosen, F., *Cognitieve organisatie van zinnen in het geheugen.* Proefschrift K.U. Leuven, 1972.

Luce, R.D., *Individual choice behavior.* New York: Wiley, 1959.

Luce, R.D., A choice theory analysis of similarity judgments. *Psychometrika, 26*, 1961, 151-163.

Lyons, J., *Introduction to theoretical linguistics.* Cambridge: Cambridge University Press, 1969.

McCawley, J.D., Concerning the base component of a transformational grammar. *Foundations of Language, 4*, 1968, 243-269.

Meerling, *Methoden en technieken van psychologisch onderzoek, dl. 2, Data-analyse en psychometrie.* Meppel (etc.): Boom, 1981.

Miller, G.A., Some psychological studies of grammar. *American Psychologist, 17*, 1962, 748-762.

Mood, A.M. & Graybill, F.A., *Introduction to the theory of statistics.* New York: McGraw-Hill, 1963.

Restle, F., *Psychology of judgment and choice.* New York: Wiley, 1961.

Restle, F., *Mathematical models in psychology. An Introduction.* Harmondsworth, Middlesex, England: Penguin Books Ltd., 1971.

Robinson, J.J., Case, category amd configuration. *Journal of Linguistics, 6*, 1969, 57-80.

Runkel, P.J. & McGrath, J., *Research on human behavior. A systematic guide to method.* New York: Holt, Rinehart and Winston, 1972

Sankoff, D. & Rousseau, P., *A method for assessing variable rule and implicational scale analysis of linguistic variation.* Paper presented at International Conference on computers in the humanities. Minneapolis, 1973.

Smeets, E., *Linguistische Intuitionen in Bezug auf die Stärke syntaktischer Relationen.* Doctoraalscriptie Instituut Duits, K.U. Nijmegen, 1982.

Spencer, N., Differences between linguists and non-linguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research, 2*, 1973, 83-98.

Spitz, J.C., De l-toets en de l'-toets, volwaardige vervangers van enkele gebruikelijke χ^2-toetsen. *Nederlands Tijdschrift voor de Psychologie, 16*, 1961, 68-88.

Tesnière, L., *Elèments de syntax structurale*. Paris: Klinsieck, 1953.

Thurstone, L.L., The prediction of choice. *Psychometrika, 10*, 1945, 237-253.

Torgerson, W.S., *Theory and methods of scaling*. New York: Wiley, 1958.

Woods, W.A., An experimental parsing system for transition network grammars. In: *Natural language processing*, R. Rustin, ed. Englewoods Cliffs, N.J.: Prentice Hall, 1973.

CURRICULUM VITAE

De schrijver van dit proefschrift is geboren te Breda op 30 mei 1942. Hij doorliep het Stedelijk Gymnasium te Maastricht en behaalde aldaar het diploma Gymnasium B in 1960. Vanaf 1961 studeerde hij Nederlandse taal- en letterkunde aan de Katholieke Universiteit van Nijmegen, deed er zijn kandidaatsexamen in 1966 en zijn doctoraalexamen in 1969. Van mei 1969 tot mei 1971 was hij als wetenschappelijk medewerker verbonden aan de afdeling Algemene Taalwetenschap en de Subfaculteit Psychologie van de Rijksuniversiteit van Groningen. In de periode van mei 1971 tot mei 1977 was hij, in dezelfde hoedanigheid, verbonden aan de vakgroep Mathematische Psychologie van de Katholieke Universiteit van Nijmegen. Sinds mei 1977 is hij werkzaam aan de Faculteit der Letteren van deze universiteit alwaar hij onderzoek verricht op het gebied van linguïstische data-analyse alsmede onderwijs en advisering inzake methodologie en statistiek verzorgt.

STELLINGEN

1. In datgene wat Chomsky karakteriseert als een dilemma tussen objectiviteit en inzicht bestaat er een "tertium datur" waarin beide doeleinden worden nagestreefd. (dit proefschrift)

   Chomsky, N., *Aspects of the theory of syntax.* Cambridge, Mass.: The M.I.T. Press, 1965 (p. 20).

2. Het is onjuist het beginsel der duidelijke gevallen (the clear case principle) los van theoretische overwegingen als bindend voor te schrijven aan de taalkundige onderzoekspraktijk. (dit proefschrift)

   Labov, W., *What is a linguistic fact ?* Lisse: The Peter de Ridder Press, 1975.

3. Afstandsmodellen lijken ongeschikt als representaties van oordelen over syntactische cohesie. (dit proefschrift)

4. Met het oog op de prominente plaats van taalintuïties in linguïstisch onderzoek verdient het aanbeveling methodologisch-linguïstisch onderzoek te stimuleren waarin de bruikbaarheid van schaal- en meettheorie voor de analyse van deze intuïties wordt onderzocht.

5. Herbert Clark heeft er terecht op gewezen dat de onjuiste beslissing om talige items als *fixed* in plaats van als *random* factoren te analyseren heeft geleid tot diverse foutieve toetsingen in psycholinguïstisch onderzoek. In zijn betoog waren de statistische grootheden min-F' en max-F' vernuftige kunstgrepen om, bij de onbeschikbaarheid van de oorspronkelijke gegevens in de betrokken literatuur, de onder- en bovengrenzen van de geïndiceerde quasi F-ratio's te reconstrueren. In situaties, evenwel, waarin men over de oorspronkelijke gegevens beschikt, kunnen quasi F-ratio's -zo nodig- zonder meer worden berekend. In die situaties heeft de vermelding van min-F' en max-F' waarden, zoals thans algemeen gebruikelijk is, weinig zin.

   Clark, H.H., The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335-359.

6. De dispersiemaat van Juilland-Chang Rodriguez, die de mate van uniforme verdeeldheid van de woordvoorkomens van een woordtype over de sub-corpora van een corpus tot uitdrukking brengt, is genormeerd op een wijze die de vergelijkbaarheid van deze maat over woorden van verschillende frequenties in de weg staat.

   Juilland, A. & Chang Rodriguez, E., *Frequency dictionary of Spanish words*. Den Haag: Mouton, 1964.

7. In de passages waarin inleidingen in de statistiek vermelden dat de steekproefvariantie met N-deling geen zuivere schatter is van de populatievariantie, wordt meestal het feit veronachtzaamd dat dit evenmin geldt voor de steekproefvariantie met (N-1)-deling bij steekproeftrekking zonder teruglegging uit eindige populaties, tenzij in dit laatste geval de populatievariantie gedefinieerd zou worden met (N_p-1)-deling (met N_p voor populatieomvang).

8. Onderzoekers die hun resultaten slechts voor de sier willen larderen met p-waarden, doen er beter aan hiertoe een generator van toevalsgetallen te gebruiken dan het omslachtige medium van statistische toepassingsprogramma's.

9. Veel onbegrip tussen mensen komt voort uit een wijdverbreide gewoonte anderen te beoordelen op die dimensies waarop men zelf het hoogste scoort onder gelijktijdige veronachtzaming van de dimensies waarop men zelf laag scoort.

10. Het dirigeren van een koor wordt bemoeilijkt door de omstandigheid dat koorzangers twee stemmen hebben. Hun ene stem is daarbij ondergeschikt aan dezelfde persoon over wie zij met hun andere stem de baas kunnen spelen.

11. Zolang de modale Nederlandse bierdrinker zijn bier snel, koud, goedkoop en in grote hoeveelheden wil consumeren, zijn de kansen van bier van hoge gisting in dit land nog maar zeer beperkt.

Stellingen behorende bij
E.D.J. Schils, Cohesion in the sentence; its use in evaluating grammars.
Diss. Nijmegen, 1983